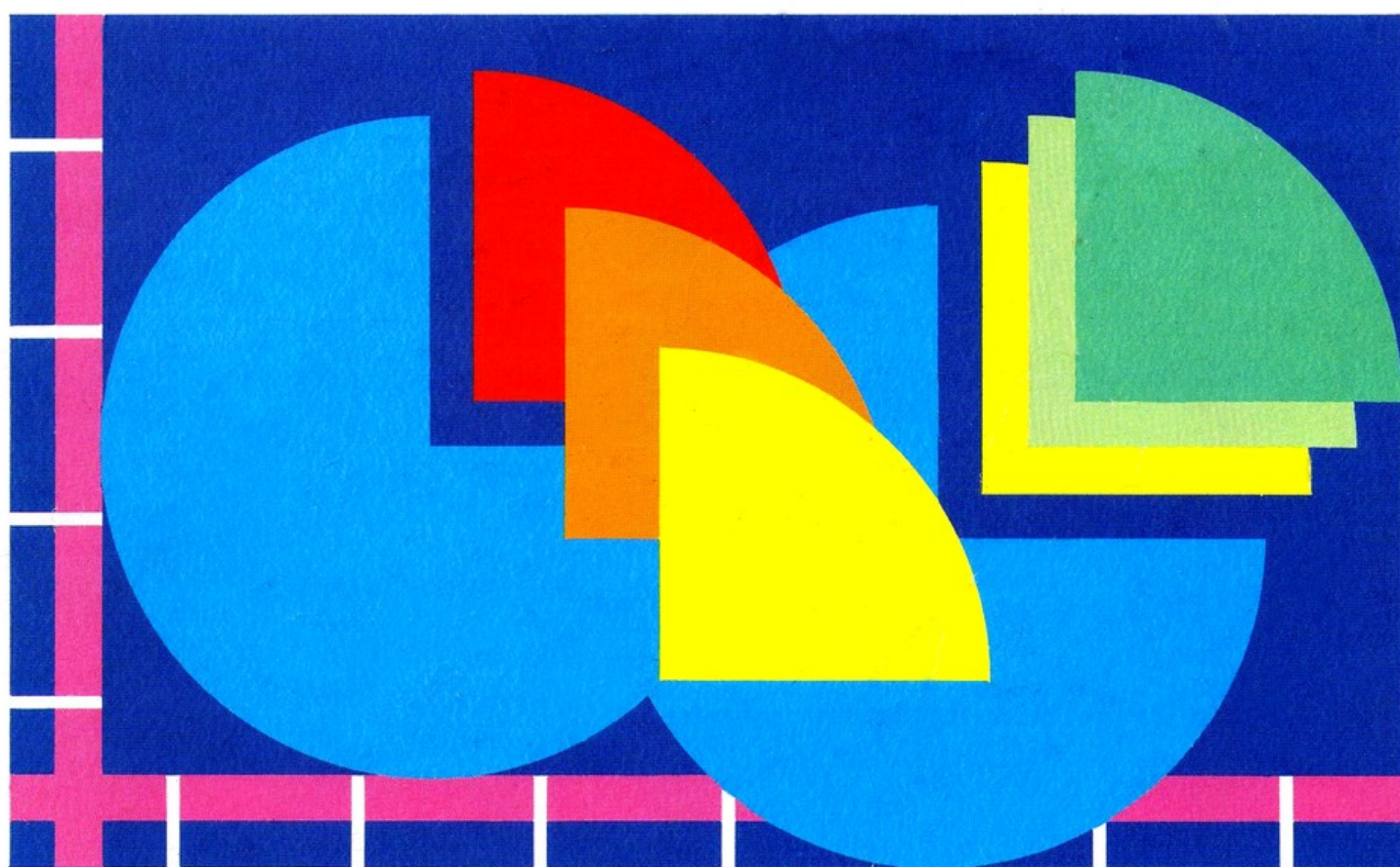
LABORATORIO SCUOLA

ELEMENTI DI BASE DI PROBABILITÀ E STATISTICA

TEORIA, ESERCIZI E SIMULAZIONI

Eugenio Rapella

ELEMENTI DI BASE DI PROBABILITÀ E STATISTICA

TEORIA, ESERCIZI E SIMULAZIONI

Eugenio Rapella

GRUPPO
EDITORIALE
JACKSON
Via Rosellini, 12
20124 Milano

CURATORI DI COLLANA: Marco Rosa-Clot, Fabio De Michele
COORDINAMENTO EDITORIALE: Emma Bennati
GRAFICA E IMPAGINAZIONE: Silvano Teruzzi
COPERTINA: Doretta Codega
FOTOCOMPOSIZIONE: Composit, Pisa

INDICE

INTRODUZIONE

Se si sfogliano i piu' recenti libri di matematica per le superiori e li si confrontano con quelli di alcuni anni fa, si nota un'aria di novita'. Per i capitoli dedicati alla probabilita' e alla statistica, si tratta addirittura di una vera rivoluzione: i contenuti, il linguaggio, le dimensioni, la collocazione e la veste tipografica hanno subito sostanziali modifiche. In effetti, e' auspicabile che uno studente della media superiore sappia qualcosa di probabilita' e di statistica (qualcosa che non si riduca a frasette del tipo ''rapporto tra casi favorevoli e casi possibili'' o ''raccolta ed elaborazione di dati) e questi temi, che gia' compaiono nei programmi delle scuole elementari e medie inferiori, vengono ora proposti, ad un livello adeguato, nel ''Programma di matematica per il biennio degli istituti ad indirizzo scientifico e tecnico''.

I primi due capitoli di questo libro sono dedicati alla probabilita': ''Elementi di calcolo delle probabilita''' e ''Problemi di calcolo delle probabilita'''. Inizialmente mi ero proposto di mescolarne il contenuto, nel tentativo di introdurre elementi di teoria solo quando questi servivano alla soluzione di un certo problema (Problema - Costruzione di un modello matematico - Acquisizione degli strumenti matematici necessari alla soluzione - Risoluzione - Simulazione).

Poiche' ne risultava un discorso piuttosto frammentario, ho deciso di sintetizzare nel primo capitolo le nozioni indispensabili e di introdurre i concetti piu' avanzati nel secondo col procedere degli esercizi.

Il lettore e' dunque invitato a dare almeno un'occhiata alla prima parte, non fosse altro che per adeguarsi alla simbologia adottata, e a scegliere quindi fra i 17 esercizi proposti nella seconda quelli che ritiene piu' interessanti.

I testi dei vari quesiti (spesso presentati come giochi) sono di im-

mediata comprensione e, nonostante nessun esercizio sia di ovvia soluzione, il ''bagaglio matematico'' di uno studente del biennio e' quanto basta per seguirne gli sviluppi (derivate ed integrali compaiono molto raramente).

Molti dei programmi Basic associati agli esercizi operano delle simulazioni il cui scopo e' quello di verificare ''sul campo'' la validita' delle conclusioni raggiunte teoricamente (ed e' con un certo piacere che si scopre come, alla lunga, le cose vadano secondo le nostre previsioni).

Il terzo capitolo e' dedicato alla statistica. Si tratta di un capitolo piuttosto breve dato che, in questa stessa collana, e' pubblicato ''Probabilita' Statistica e Termodinamica'', che contiene, su questo argomento, molto materiale (e programmi).

I temi proposti nel terzo capitolo dovrebbero bastare per uno studente del biennio. La presentazione di argomenti piu' comples-si richiederebbe conoscenze matematiche troppo avanzate e, in mancanza di queste, apparirebbe poco convincente.

Questo libro ha come destinatari sia studenti che insegnanti; le parti piu' formali sono spesso inserite per completezza e si rivolgono principalmente agli insegnanti. Di conseguenza, lo studente potra' trovare un po' difficili alcuni paragrafi del primo capitolo: senza preoccuparsene eccessivamente, cerchi di assimilare i concetti e di familiarizzare con le idee fondamentali. Potra' rivedere a suo tempo le parti piu' formali, eventualmente con l'aiuto dell'insegnante.

Sono molto grato a Sergio Casiraghi, che ha curato la versione MS DOS dei programmi, e ad Antonella Vitalini che ha pazientemente letto una prima stesura del testo.

C'e' una persona che voglio ringraziare in modo particolare: Nicolo' Pintacuda, docente di calcolo delle probabilita' presso l'universita' di Pavia. Ho avuto la fortuna di assistere alle sue lezioni durante gli anni dell'universita' e gli sono debitore non solo per i consigli e suggerimenti relativi al contenuto di questo libro, ma, soprattutto, per aver suscitato il mio interesse per uno dei piu' affascinanti capitoli della matematica.

Eugenio Rapella

CAPITOLO 1

ELEMENTI DI CALCOLO DELLE PROBABILITÀ

In questo primo capitolo vengono riassunte le nozioni base di cui si fara' ampiamente uso nel secondo, dedicato agli esercizi.

A chi desidera approfondire questa parte consiglio i testi [14] e [15] citati nella bibliografia, si tratta di libri didatticamente pregevoli e di piacevole lettura.

Il calcolo delle probabilita' viene egregiamente trattato in diversi libri di testo per le superiori; fra quelli a mia conoscenza meritano una particolare segnalazione [4], [8], [19] in bibliografia.

1.1 LA PROBABILITA'

Fornire una definizione elementare di probabilita' non e' semplice e, di norma, i libri di testo procedono ''per piccoli passi'' distinguendo diverse formulazioni.

In questo paragrafo inizieremo con l'impostazione assiomatica, dove la probabilita' viene definita attraverso alcune proprieta' di cui deve ragionevolmente godere. Successivamente si vedra' come, a seconda delle informazioni sul fenomeno casuale, si giustifichino la concezione ''classica'', quella ''frequentista'' e quella ''soggettiva''.

Giungeremo alla definizione attraverso una serie di osservazioni:

1 - E' facile trovare esempi di fatti che possono accadere oppure no; li chiameremo EVENTI ALEATORI. Li identificheremo con degli insiemi, sottoinsiemi di un insieme piu' Ω vasto che comprende tutte le eventualita'; Ω sara' ''l'EVENTO CERTO''.

Un evento aleatorio sara' indicato con una lettera maiuscola e, quando necessario, ne sara' data una descrizione, tra parentesi graf-

fe, con una frase o con un elenco di eventi elementari che caratterizzano la sua realizzazione.

Esempio

Consideriamo la situazione in cui viene lanciato un normale dado a sei facce. L'insieme Ω sara' l'uscita di uno dei valori da 1 a 6
$\Omega = \{1,2,3,4,5,6\}$

Esempi di eventi aleatori sono:

A = {uscita di un valore minore di 5} = {1,2,3,4}
B = {uscita di un valore tra 3 e 5} = {3,4,5}
Come si nota, $A \subseteq \Omega$ e $B \subseteq \Omega$.

Unione ed Intersezione di eventi

2 - Se A e B sono due eventi, indicheremo con:

$A \cup B$ l'evento che ha luogo se si verifica A oppure si verifica B (o entrambi);

$A \cap B$ l'evento che ha luogo se si verifica sia A che B;

$\mathscr{C}(A)$ l'evento che si verifica quando non si verifica A (l'evento contrario). Corrisponde all'insieme complementare di A rispetto a Ω.

Riprendendo l'esempio precedente, avremo:
$A \cup B = \{1,2,3,4,5\}$
$A \cap B = \{3,4\}$
$\mathscr{C}(A) = \{5,6\}$
$\mathscr{C}(B) = \{1,2,6\}$

3 - L'insieme vuoto $\varnothing$ rappresentera' un EVENTO IMPOSSIBILE. Nel lancio del dado, l'evento
C = {uscita di un valore maggiore di 6}
e' un esempio di evento impossibile: $C = \varnothing$.

Eventi incompatibili

4 - Diremo INCOMPATIBILI due eventi A e B per i quali e' $A \cap B = \varnothing$ (e' impossibile che si verifichino entrambi). Nel lancio del dado, posto $X = \{1,2\}$, $Y = \{4,5,6\}$ si ha $X \cap Y = \varnothing$.

Qualunque sia l'evento A e' facile verificare che:
$A \cup A = A$;
$A \cap A = A$;
$A \cup \varnothing = A$;
$A \cap \varnothing = \varnothing$;
$A \cup \mathscr{C}(A) = \Omega$;
$A \cap \mathscr{C}(A) = \varnothing$

5 - Dire che A e' un sottoinsieme di B ($A \subseteq B$) equivarra' a dire che, quando si verifica A, si verifica anche B, ovvero che ''A implica B''. Nel caso del dado cio' avviene, per esempio, per
$A = \{2\}$

e

B = {uscita di un numero pari}.

6 - Consideriamo una scatola che contiene 100 palline di ugual peso e dimensioni: 99 sono nere, una e' bianca.

Viene estratta una pallina, sia

A = {viene estratta una pallina nera}

B = {viene estratta una pallina bianca} = $\mathscr{C}$(A).

A e B non sono ne' certi ne' impossibili, ma l'evento A si verifica ''piu' facilmente'' dell'evento B.

Probabilità

Volendo quantificare diversi gradi di incertezza, possiamo pensare di associare ad un evento A un numero reale P(A) - PROBABILITA' DI A - che esprima una misura di ''quanto'' l'evento A possa accadere. Diremo MISURABILI quegli eventi a cui e' associato un valore di probabilita'.

7 - L'insieme di tutti gli eventi misurabili costituisce una famiglia di $\mathscr{A}$ sottoinsiemi di Ω.

Non e' indispensabile che $\mathscr{A}$ sia formata da tutti i sottoinsiemi di Ω (insieme delle parti $\mathscr{P}(\Omega)$); e' comunque ragionevole postulare che, se A e B $\in \mathscr{A}$, anche $A \cup B \in \mathscr{A}$; $A \cap B \in \mathscr{A}$; $\mathscr{C}(A) \in \mathscr{A}$.

In altri termini, se agli eventi A e B e' associato un valore di probabilita', avra' senso misurare la probabilita' che accadano entrambi o uno almeno dei due ecc.

Impostazione assiomatica

8 - Se richiediamo che:

- $\forall A \in \mathscr{A}$ sia $P(A) \geqslant 0$
(la probabilita' di un evento e' non negativa)

- $P(\Omega) = 1$
(la probabilita' dell'evento certo e' 1)

- $\forall A \in \mathscr{A}$, $\forall B \in \mathscr{A}$ con $A \cap B = \emptyset$ sia $P(A \cup B) = P(A) + P(B)$
(postulato di additivita')

la probabilita' ha tutti i requisiti per essere una misura sensata dell'incertezza di un evento. Dagli assiomi discendono infatti diverse proprieta' che era logico attendersi; eccone alcune:

$P(\emptyset) = 0$: l'evento impossibile ha probabilita' nulla (infatti $\Omega = \emptyset \cup \Omega$ disgiunti quindi $P(\Omega) = P(\emptyset) + P(\Omega)$ ...)

$A \subseteq B \Rightarrow P(A) \leqslant P(B)$: se un evento ne implica un altro, la probabilita' del primo non supera quella del secondo (infatti, indi-

cato con B\A l'evento che si verifica quando si verifica B e non si verifica A, si ha:

$B = A \cup (B \setminus A)$ disgiunti quindi
$P(B) = P(A) + P(B \setminus A) \geqslant P(A)$)

Qualunque sia l'evento A e' $\emptyset \subseteq A \subseteq \Omega$ per cui $0 \leqslant P(A) \leqslant 1$: la probabilita' di un evento e' un numero compreso tra 0 e 1

$P(\mathscr{C}(A)) = 1 - P(A)$ (infatti $\Omega = A \cup \mathscr{C}(A)$ disgiunti quindi $P(\Omega) = P(A) + P(\mathscr{C}(A))$ ...)

Riassumendo, fissato l'insieme Ω di tutte le eventualita' e una famiglia $\mathscr{A}$ di sottoinsiemi di Ω chiusa per unione (*), intersezione (*), passaggio al complementare; la probabilita' viene definita come una applicazione da $\mathscr{A}$ in [0,1] tale che:

- $P(\Omega) = 1$
- $\forall$ A,B $\in \mathscr{A}$ con $A \cap B = \emptyset$
sia $P(A \cup B) = P(A) + P(B)$ (*)

[In generale le proprieta' (*) vengono estese ad una infinita' numerabile di eventi, ovvero:

Se $A_1, A_2, \ldots, A_n, \ldots$ e' una successione di elementi di $\mathscr{A}$ si ha $\cup_i A_i \in \mathscr{A}$ e $\cap_i A_i \in \mathscr{A}$; se gli eventi A_i sono a due a due disgiunti ($A_i \cap A_j = \emptyset$ per $i \neq j$), la serie $\Sigma_n P(A_n)$ converge e la sua somma e' $P(\cup_n A_n)$.]

La definizione non dice dunque come misurare la probabilita' di un generico evento (ne' ci si poteva aspettare qualcosa di diverso); val la pena comunque di sottolineare come la probabilita' di un evento possa cambiare al variare delle informazioni in nostro possesso sul fenomeno in discussione.

Se
A = {La squadra vincitrice del campionato italiano di calcio del 1975 e' la Juventus},
alla domanda
''Qual e' la probabilita' di A'' ?
posso rispondere P(A) = 1, visto che ho appena controllato su un'enciclopedia (di cui mi fido). Se la stessa domanda mi fosse stata rivolta mezz'ora fa, avrei risposto con un valore minore di uno (dato che non mi interesso di calcio) e cosi', a maggior ragione, se mi fosse stata rivolta nel 1974.

Ritorniamo ad un esempio descritto in precedenza: un normale dado a sei facce viene lanciato; ci si chiede la probabilita' dell'evento A = {uscita di un valore minore di 5}.

Costruiamo Ω e $\mathscr{A}$:

$\Omega = \{1,2,3,4,5,6\}$; $\mathscr{A} = \mathscr{P}(\Omega)$.

Consideriamo i sei eventi elementari $B_i = \{i\}$, $i = 1,2,..,6$ e notiamo che:

$\Omega = B_1 \cup B_2 \cup B_3 \cup B_4 \cup B_5 \cup B_6$
(e' certo che uscira' un numero tra 1 e 6)

$B_i \cap B_j = \emptyset$ se $i \neq j$
(l'uscita di un certo numero esclude l'uscita di uno degli altri cinque)

Per il postulato di additivita' e':

$$P(\Omega) = P(B_1) + P(B_2) + P(B_3) + P(B_4) + P(B_5) + P(B_6).$$

Si noti come, nella costruzione di Ω, abbiamo gia' tenuto conto di una informazione: il dado ha sei facce contrassegnate dai valori da 1 a 6. Abbiamo un'ulteriore informazione: si tratta di un ''normale'' dado; con questo intendiamo che l'uscita di un certo numero non e' ne' ''piu' facile'' ne' ''piu' difficile'' dell'uscita di un altro.

Cio' significa che, per un dado di questo tipo che chiameremo ''regolare'', le probabilita' degli eventi B_i $(i = 1,2,..,6)$ sono uguali tra loro. Se indichiamo con ''x'' il loro comune valore, otteniamo:

$$P(\Omega) = x + x + x + x + x + x$$

ovvero $6x = 1$ e quindi $x = 1/6$.

A questo punto, per calcolare P(A), bastera' notare che A e' l'unione dei quattro eventi incompatibili B_1, B_2, B_3, B_4 .

Applicando nuovamente il postulato di additivita', avremo:

$$P(A) = (1/6) + (1/6) + (1/6) + (1/6) = 2/3 .$$

Sfruttando dunque in modo opportuno le informazioni, si e' giunti alla valutazione effettiva della probabilita' di un evento.

In questo esempio, le condizioni che hanno reso possibile questa operazione sono due:

il fatto che l'insieme Ω fosse costituito da un numero finito di eventi elementari
l'aver assegnato ad ogni evento elementare la stessa probabilita'.

Impostazione classica

In situazioni di questo tipo, la probabilita' di un evento A risulta pari al rapporto tra il numero di elementi di A e il numero di elementi di Ω. Nell'impostazione classica, la probabilita' di un evento viene definita proprio in questi termini:

> ''La probabilita' di un evento A e' il rapporto tra il numero m di casi favorevoli al verificarsi di A e il numero n di casi possibili, giudicati ugualmente possibili:
> $P(A) = m/n$ ''

Questa definizione e' stata giustamente criticata, sia perche' puo' essere impiegata solo in casi particolari (non sempre sussistono le condizioni 1 e 2), sia perche' non viene indicato quando i ''casi possibili'' siano da giudicarsi ''ugualmente possibili'' (ne' si puo' dire che gli eventi elementari sono ''ugualmente possibili'' se hanno la stessa probabilita', dato che la probabilita' e' cio' che si vuol definire).

Non e' neppure detto che una stima sensata della probabilita' di un evento sia possibile solo attraverso la definizione classica: negli esercizi 8 e 9 del secondo capitolo si valutano delle probabilita' in situazioni in cui i casi favorevoli e quelli possibili non sono in numero finito.

Per quanto riguarda la costruzione di Ω e $\mathscr{A}$, si consideri il lancio di una moneta regolare. Posto

$T = \{$uscita della faccia ''testa''$\}$;
$C = \{$uscita della faccia ''croce''$\}$,

costruiamo

$\Omega = \{T,C\}$ e
$\mathscr{A} = \mathscr{P}(\Omega) = \{\varnothing,\Omega,T,C\}$.

Ragionando come nell'esempio del dado, si avra' $P(T) = P(C) = 1/2$.

Supponiamo ora di voler trattare il problema del lancio di una moneta regolare, ma di disporre di un dado regolare. Possiamo utilizzare il dado in luogo della moneta costruendo Ω e $\mathscr{A}$ in modo opportuno; bastera' considerare i due eventi

$A = \{$uscita di un numero pari$\}$
$B = \{$uscita di un numero dispari$\}$

e porre

$\Omega = \{A,B\}$, $\mathscr{A} = \mathscr{P}(\Omega)$.

Si avra' P(A) = P(B) = 1/2 ottenendo un modello identico al precedente. Con cio' si vede come, per uno stesso esperimento aleatorio, la coppia $(\Omega, \mathscr{A})$ possa essere costruita in modi differenti per soddisfare esigenze diverse.

Consideriamo la seguente situazione:

> ''Una scatola contiene palline bianche e nere, indistinguibili al tatto, per un totale di 100 palline. Si vuole estrarre una pallina; sia
> A = {estrazione di una pallina bianca}.

Ci proponiamo di stimare P(A). L'impostazione classica non e' di grande aiuto dato che non e' noto il numero di ''casi favorevoli'' all'evento A; P(A) puo' essere un qualsiasi valore del tipo k/100 dove k rappresenta un intero tra 0 e 100.

Integriamo il testo con una nuova informazione:

> ''Dalla scatola sono state in precedenza eseguite 1000 estrazioni rimettendo ogni volta nella scatola la pallina estratta: 950 volte e' stata estratta una pallina bianca; 50 volte una pallina nera''.

Se chiamiamo ''FREQUENZA DI A'' il rapporto tra il numero di prove nelle quali A si e' verificato e il numero delle prove eseguite, potremo dire che la frequenza di A e' pari a 950/1000 = 0.95 .

L'informazione supplementare ha cambiato la situazione e l'ipotesi che A si verifichi piu' facilmente di $\mathscr{C}$(A) ne esce rafforzata.

Impostazione frequentista

Come si vedra' nel paragrafo 1.6, sotto certe ipotesi, la frequenza puo' essere considerata una buona stima della probabilita'; nell'impostazione frequentista, la probabilita' di un evento viene addirittura identificata con la frequenza relativa ad un numero di prove, eseguite nelle stesse condizioni, ritenuto sufficientemente elevato.

Senza entrare nel merito delle critiche che si possono muovere a questa definizione (che, fra l'altro, renderebbe misurabili solo quegli eventi per i quali e' concepibile una ripetizione dell'esperimento aleatorio in condizioni immutate), si potra' accettare la frequenza come stima della probabilita'. In altri termini, nell'esempio della scatola con le 100 palline, non diremo che P(A) e' 0.95, ma che il valore 0.95 e' quanto di meglio abbiamo per stimare P(A).

"Probabilità statistiche"

Il calcolo delle probabilita' e la statistica forniscono gli strumenti per valutare i rischi di approssimazioni di questo tipo ed e' per questo che ''PROBABILITA' STATISTICHE'' vengono comunemente utilizzate quando si dispone di dati statistici relativi a fenomeni che, in linea di massima, si ripresentano in condizioni analoghe (probabilita' di sopravvivenza, probabilita' di incidenti, probabilita' di particolari situazioni meteorologiche ecc.).

Abbiamo visto che la valutazione della probabilita' di un evento applicando la definizione classica o frequentista e' possibile in ambiti molto ristretti. Nonostante cio', ciascuno di noi e' chiamato ogni giorno a prendere delle decisioni in condizioni di incertezza e, piu' o meno consciamente, a valutare delle probabilita'.

Supponiamo che il pugile X, detentore del titolo mondiale di una certa categoria, stia per incontrare lo sfidante Y; sia

A = {dopo l'incontro X sara' riconfermato campione del mondo}.

La probabilita' di A non si presta ad essere valutata ne' secondo lo schema classico (non c'e' ragione di pensare che A e $\mathscr{C}$(A) siano equiprobabili), ne' secondo quello frequentista (i due pugili potrebbero non essersi mai incontrati e, anche in caso contrario, le ''prove'' non sarebbero ne' sufficientemente numerose ne' ''eseguite nelle stesse condizioni'').

C'e' di piu'; l'insieme delle informazioni relative a questa situazione e' talmente articolato che due persone diverse possono benissimo attribuire ad A diverse valutazioni di probabilita': c'e' chi scommette su X, c'e' chi scommette su Y.

Nell'impostazione soggettiva, la probabilita' di un evento A viene proprio definita ipotizzando una scommessa:

Impostazione soggettiva

''La probabilita' di un evento A e' quella somma P(A) che:
1 accetterei preventivamente di pagare per ricevere 1 lira al verificarsi di A
e, coerentemente,
2 accetterei preventivamente di ricevere, impegnandomi a pagare 1 lira se A si verifica.''

La disponibilita' ad accettare indifferentemente una o l'altra delle scommesse, fa' si' che $\forall$ A sia $0 \leq P(A) \leq 1$ (se fosse $P(A) > 1$, la prima scommessa risulterebbe comunque svantaggiosa; per $P(A) < 0$, lo sarebbe comunque la seconda).

Vediamo come valutare la probabilita' di un evento ritenuto certo (Ω). Con la scommessa (1) sono sicuro di ricevere una lira: per giocare sono disposto a pagare una qualsiasi somma $x \leq 1$. Accettan-

do la scommessa (2) dovro' certamente pagare una lira; giocare mi conviene solo se mi viene preventivamente versata una somma $x \geq 1$. Si ha dunque $x \leq 1$ e $x \geq 1$ per cui $P(\Omega) = 1$. In modo analogo si puo' verificare che $P(\emptyset) = 0$.

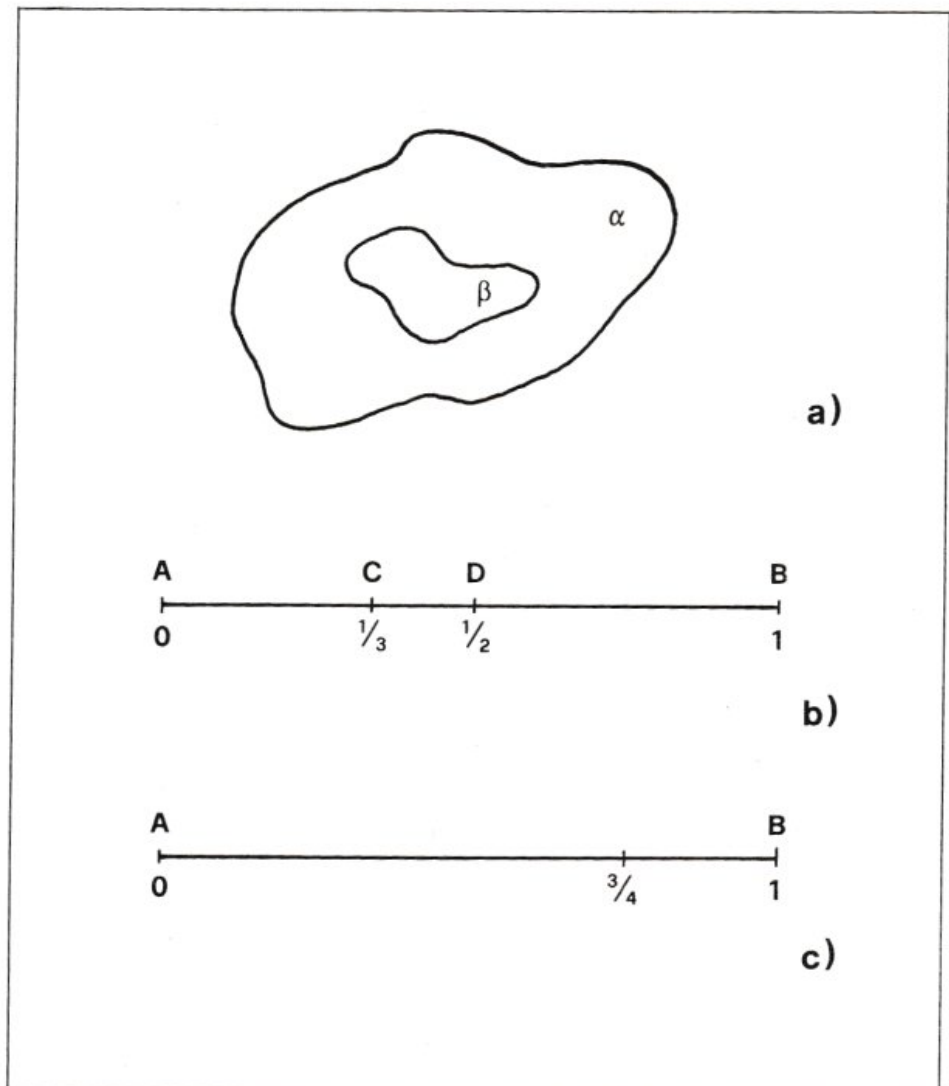
L'impostazione soggettiva formalizza la concezione di probabilita' che si ha comunemente, compresa quella possibilita' di diverse valutazioni da parte di diversi individui che, di fatto, si verifica per molti eventi.

Quest'ultima definizione si puo' applicare praticamente a qualsiasi evento.

Concludiamo questo paragrafo con due esempi:

Probabilità "geometriche"

Es.1 Un punto X e' scelto a caso (nel senso che ogni punto ha la stessa probabilita' di essere scelto) nella regione α (fig.1a); qual e' la probabilita' che X appartenga alla regione β?

Fig. 1 *Esempi di probabilità geometriche*

I ''casi possibili'' sono gli infiniti punti di α, quelli ''favorevoli'' gli infiniti punti di β e la definizione classica non puo' essere utilizzata. In un certo senso, l'area di una data regione ne misura il numero di punti ed e' ragionevole porre $P[X \epsilon \beta] = \text{Area}(\beta)/\text{Area}(\alpha)$.

Es.2 Si sceglie a caso un numero reale x compreso tra 0 e 1. Qual e' la probabilita' che sia $1/3 \leqslant x \leqslant 1/2$? E che sia $x = 3/4$?

Ragionando come nell'esempio precedente (vedi fig. 1b e 1c), avremo:

$$P[1/3 \leqslant x \leqslant 1/2] = \frac{\overline{CD}}{\overline{AB}} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

mentre $P[x = 3/4] = 0$.

Si noti come l'evento $(x = 3/4)$, pur non essendo rigorosamente impossibile, abbia probabilita' nulla (si usa dire che l'evento e' ''quasi impossibile'' e il suo complementare ''quasi certo''). La cosa non deve sorprendere piu' di tanto; come fa notare F.Speranza ([18], pag.60), quando si lavora su insiemi infiniti, si presentano delle ''stranezze'' e il calcolo delle probabilita' non fa eccezione.

Probabilita' valutate attraverso modelli geometrici saranno utilizzate negli esercizi 8 e 9 del 2' capitolo. Naturalmente l'estrazione di un elemento da un insieme che ne contiene infiniti non e' attuabile praticamente e le simulazioni su calcolatore forniranno una realizzazione approssimata dell'esperimento. Cosi', se viene generato un valore pseudocasuale (vedi paragrafo 1.8) dell'intervallo (0,1), la probabilita che x sia uguale a 3/4 risulta essere maggiore di zero poiche', nel calcolatore, il numero di casi possibili e' finito.

1.2 PROBABILITA' CONDIZIONALE

Dati due eventi A e B con $P(B) > 0$, chiamiamo PROBABILITA' CONDIZIONALE DI A DATO B (o, piu' semplicemente, PROBABILITA' DI A DATO B) la probabilita' che si verifichi l'evento A nell'ipotesi che B si sia verificato.

Esempio. Nel lancio di un dado regolare sia:

$A = \{1,2,3\}$
$B = \{2,3,4,5,6\}$

La probabilita' di A dato B, che indicheremo con P(A/B), e' la probabilita' che si verifichi A sapendo che e' uscito un numero maggiore di uno.

Per effetto della nuova informazione, i casi possibili si riducono

a 5 (2,3,4,5,6) di cui solo 2 favorevoli al verificarsi di A (2,3). Si ha dunque $P(A/B) = 2/5$.

L'ipotesi che B si sia verificato, restringe l'insieme delle eventualita' ai soli elementi di B; contribuiranno alla probabilita' di A solo quegli elementi di A compatibili con B. E' logico quindi porre

$$P(A/B) = P(A \cap B)/P(B)$$

quantita' che misura l'incidenza relativa di A in B.

Per un evento B fissato, la probabilita' condizionale soddisfa agli assiomi della probabilita'; sono di immediata verifica le seguenti proprieta':

$\forall$ A e' $P(A/A) = 1$
(E' certo che A si verifichi, se ci si pone nell'ottica che si sia verificato)

Se $A \cap B = \emptyset$ e' $P(A/B) = 0$
(Poiche' A e' incompatibile con B, e' impossibile che A si verifichi se supponiamo verificato B)

$\forall$ A e' $P(A/\Omega) = P(A)$.

Sia

$H_1, H_2, .., H_n$ una partizione di Ω
($H_i \neq \emptyset \ \forall i$; $H_i \cap H_j = \emptyset$ se $i \neq j$; $H_1 \cup H_2 \cup .. \cup H_n = \Omega$);
sia $A \subseteq \Omega$ e $H_i \in \mathscr{A} \forall i$, $A \in \mathscr{A}$.

Poiche':

$$A = (A \cap H_1) \cup (A \cap H_2) \cup ... \cup (A \cap H_n)$$
$A \cap H_i$ e' incompatibile con $A \cap H_j$ per $i \neq j$;
e' possibile applicare il postulato di additivita':

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + .. + P(A \cap H_n).$$

Dalla definizione di probabilita' condizionale si ha che $P(A \cap H_i) = P(A/H_i)P(H_i)$ per cui

$$P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + .. + P(A/H_n)P(H_n) \ .$$

Questa uguaglianza e' utilissima perche', molto spesso, risulta piu' semplice il calcolo delle diverse probabilita' condizionali al secondo membro piuttosto che il computo diretto della probabilita'

Fig. 2 *Matrice per la soluzione del problema delle 7 palline*

al primo membro; si confrontino i due metodi di soluzione del seguente problema:

Problema

''Una scatola contiene 7 palline: 4 rosse, 3 bianche. Ne vengono estratte 2; calcolare la probabilita' dell'evento

A = {le palline estratte sono di colore diverso}

Soluzione 1

Pensiamo di numerare le palline rosse: R1, R2, R3, R4 e quelle bianche: B1, B2, B3.

Nella matrice di fig.2

- le colonne corrispondono ai risultati della prima estrazione
- le righe corrispondono ai risultati della seconda estrazione.

Ogni elemento della matrice, ad eccezione di quelli della diagonale principale (la stessa pallina non puo' essere estratta due volte), corrisponde ad uno dei casi possibili. Le caselle contrassegnate con ''X'' identificano i casi favorevoli.

Procedendo al conteggio, si ottiene:
''casi possibili'' : 42
''casi favorevoli'': 24

P(A) = 24/42 = 4/7

Soluzione 2

Sia H = {rossa alla 1' estrazione}

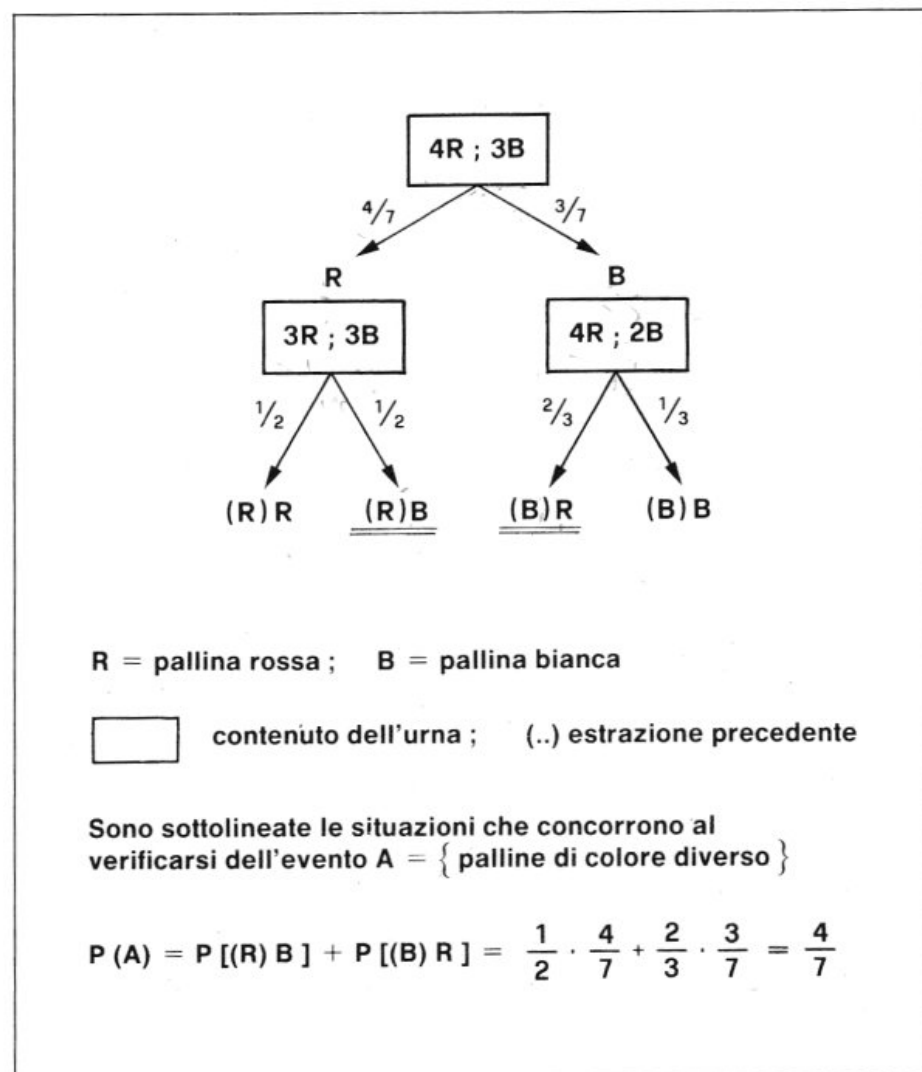
K = {bianca alla 1' estrazione}
e' $H \cap K = \emptyset$ e $H \cup K = \Omega$;
la coppia H,K costituisce una partizione di Ω per cui
$P(A) = P(A/H)P(H) + P(A/K)P(K)$

$P(H) = 4/7$. Supponiamo che H si sia verificato, l'urna conterra' 3 rosse 3 bianche. Perche' si realizzi A si deve ora estrarre una bianca: $P(A/H) = 1/2$.

$P(K) = 3/7$. Supposto verificato K, l'urna contiene 4 rosse e 2 bianche. ''A'' si verifica se ora viene estratta una rossa, per cui: $P(A/K) = 4/6 = 2/3$.

In definitiva
$P(A) = (4/7)(1/2) + (3/7)(2/3) = 4/7$

Il calcolo diviene molto semplice con l'ausilio di un GRAFO (fig.3): ad ogni nodo e' associata una situazione e da esso si staccano i rami delle situazioni ottenibili da quella relativa al nodo.

Fig. 3 *Grafo per la soluzione del problema delle 7 palline*

A fianco di ogni ramo e' espressa la probabilita' del verificarsi dell'evento ''di arrivo'' a condizione che si sia verificato quello ''di partenza''. La probabilita' di un evento X si ottiene moltiplicando le probabilita' dei rami che conducono a situazioni ''favorevoli a X'', sommando poi i vari contributi.

1.3 INDIPENDENZA

Dati due eventi A e B, puo' essere $P(A/B) < P(A)$, $P(A/B) > P(A)$ o $P(A/B) = P(A)$.

L'ultimo caso e' particolarmente interessante: il fatto di supporre B verificato non ha cambiato la probabilita' di A. Quando $P(A/B) = P(A)$, diremo che l'evento A e' stocasticamente INDIPENDENTE dall'evento B.

Eventi indipendenti

Esempio 1

Vengono lanciati due dadi regolari, uno rosso e uno bianco. Sia

A = {uscita del 6 sul dado rosso}
B = {uscita del 6 sul dado bianco}

e' $P(A) = P(B) = 1/6$, $P(A \cap B) = 1/36$.

$P(A/B) = P(A \cap B)/P(B) = (1/36)/(1/6) = 1/6$

quindi $P(A/B) = P(A)$: la probabilita' che esca 6 sul dado rosso non viene alterata dall'informazione che e' uscito un 6 su quello bianco; A e' indipendente da B.

Esempio 2

Lancio di un dado regolare, sia

$A = \{1,2\}$
$B = \{2,4,6\}$.

E' $P(A) = 1/3$; $P(A/B) = P(A \cap B)/P(B) = (1/6)/(1/2) = 1/3$.

A e' indipendente da B (anche se i due eventi si riferiscono allo stesso esperimento aleatorio).

Se A e' indipendente da B, anche B risulta indipendente da A, infatti

$$P(B/A) = P(B \cap A)/P(A) = P(A/B)P(B)/P(A) =$$
$$= P(A)P(B)/P(A) = P(B).$$

Data la simmetria, bastera' dire che gli eventi A e B sono indi-

pendenti (uno dall'altro) e la situazione di indipendenza potra' essere caratterizzata dalla proprieta':

A e B sono indipendenti $\Leftrightarrow$ $P(A \cap B) = P(A)P(B)$.

Una famiglia A_1 , A_2 ,..., A_n di eventi si dira' COMPLETAMENTE INDIPENDENTE se, comunque se ne scelgano k, la probabilita' che si verifichino tutti e' pari al prodotto delle probabilita' di ciascuno di essi. In simboli:

$P(A_{i_1} \cap A_{i_2} \cap .. \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})..P(A_{i_k})$

per qualsiasi scelta degli indici i_j .

1.4 VARIABILI ALEATORIE

Esempio 1

Un giocatore lancia un dado regolare e riceve un numero di dollari pari al doppio del punteggio ottenuto. Detta X la somma guadagnata dal giocatore, X puo' assumere uno dei sei valori 2, 4, 6, 8, 10, 12 e, piu' precisamente:

X vale 2 se e' uscito il numero 1;
X vale 4 se e' uscito il numero 2;
ecc.

''X'' e' dunque una variabile che ''dipende dal caso'' nel senso che assume certi valori al verificarsi di certi eventi.

Gli esempi di variabili di questo tipo sono innumerevoli e il loro studio e' di fondamentale importanza nel calcolo delle probabilita'; ci limitiamo a considerare il caso in cui la variabile assume un numero finito o una infinita' numerabile di valori reali: VARIABILI ALEATORIE REALI DISCRETE.

Fissato l'insieme Ω delle eventualita', sia $H_1, H_2, .., H_n, ..$ una partizione finita o numerabile costituita da sottoinsiemi misurabili di Ω. Chiameremo VARIABILE ALEATORIA una funzione X che assume il valore reale x_i su H_i.

Poiche' gli eventi H_i costituiscono una partizione di Ω, posto $p_i = P(H_i)$, si avra' $\Sigma_i p_i = 1$.

Nell'esempio 1 e' $\Omega = \{1,2,3,4,5,6\}$
$H_i = \{i\}$; $p_i = 1/6$; $x_i = 2i$ per $i = 1,2,..,6$

Legge di una variabile aleatoria

L'insieme delle coppie (x_i, p_i) definisce la LEGGE di una variabile aleatoria discreta; variabili aleatorie aventi la stessa legge si dicono EQUIDISTRIBUITE.

Esempio 2

Lancio un dado regolare; se esce un numero pari vinco 100 lire; se esce un numero dispari perdo 100 lire. Sia Y = ''somma vinta''; Y e' una variabile aleatoria di legge:

y_i	+100	-100
p_i	1/2	1/2

Esempio 3

Lancio una moneta regolare; vinco 100 lire se esce ''testa'', perdo 100 lire se esce ''croce''. Sia Z = ''somma vinta''; la legge di Z e'

z_i	+100	-100
p_i	1/2	1/2

Le variabili Y e Z sono equidistribuite. Come si nota, la legge fornisce, molto spesso, un'adeguata descrizione della variabile aleatoria senza che sia necessaria una costruzione dettagliata di Ω e della relativa partizione (il gioco proposto negli esempi 2 e 3 e' fondamentalmente lo stesso: con probabilita' 0.5 si vincono 100 lire, con probabilita' 0.5 si perdono 100 lire).

Media di una variabile aleatoria

Se X e' una variabile aleatoria di legge (x_i, p_i), si definisce MEDIA (o SPERANZA MATEMATICA o VALORE ATTESO) di X il valore

$$E[X] = \Sigma_i p_i x_i$$

(nel caso di una infinita' numerabile di x_i, la media e' definita solo se la serie converge).

La media fornisce una stima globale di una variabile aleatoria: ogni valore x_i contribuisce al computo di E[X] con un peso pari alla probabilita' di essere assunto.

Per quanto riguarda gli esempi 1, 2 e 3, avremo:

$$E[X] = (1/6)(2+4+6+8+10+12) = 7$$
$$E[Y] = 0 \; ; \; E[Z] = 0.$$

Variabili aleatorie equidistribuite hanno evidentemente la stessa media, ma l'uguaglianza delle medie non implica, in generale, l'uguaglianza delle leggi (se, nell'esempio 3, si sostituiscono i valori +100 e -100 con +1000 e -1000, si ottiene una variabile aleatoria di ugual media e legge diversa).

Varianza di una variabile aleatoria

I valori x_i di una variabile aleatoria X possono essere piu' o meno concentrati attorno al valore medio $\mu = E[X]$; una misura del ''grado di dispersione'' e' fornita dalla VARIANZA di X:

$$\sigma^2(X) = E[(X - \mu)^2] = \sum_i (x_i - \mu)^2 p_i .$$

$\sigma(X)$, radice quadrata della varianza, e' detta SCARTO QUADRATICO MEDIO (o DEVIAZIONE STANDARD) della variabile aleatoria X.

Esempio 4
Sia X di legge:

x_i	4	10	20
p_i	1/4	1/2	1/4.

Calcolare E[X], $\sigma^2(X)$.
Avremo:

$\mu = E[X] = 4(1/4) + 10(1/2) + 20(1/4) = 11$

$\sigma^2 (X) = (4\text{-}11)^2(1/4) + (10\text{-}11)^2(1/2) + (20\text{-}11)^2(1/4) =$
$= 12.25 + 0.5 + 20.25 = 33$

$\sigma(X) = \sqrt{33} = 5.74$

Giochi

Se, come negli esempi 1, 2 e 3, i valori di una variabile aleatoria X rappresentano i guadagni (positivi o negativi) associati al realizzarsi di certi eventi, E[X] rappresenta il ''guadagno medio teorico'' ovvero la somma che un giocatore puo' aspettarsi di guadagnare mediamente per ogni prova eseguita (se le prove vengono eseguite nelle stesse condizioni e il loro numero e' sufficientemente elevato).

Il gioco e' detto:

VANTAGGIOSO se $E[X] > 0$
SVANTAGGIOSO se $E[X] < 0$
EQUO se $E[X] = 0$

Il gioco descritto nell'esempio 1 e' vantaggioso (in effetti si puo' soltanto vincere), quello dell'esempio 2 e' equo.

Se due variabili aleatorie hanno la stessa media, quella con varianza maggiore e' ''piu' aleatoria'': lo scarto dal valore atteso e', mediamente, piu' grande.

Dovendo scegliere tra due investimenti

Invest.1 X: guadagno -100 +200
con prob. 1/2 1/2

Invest.2 Y: guadagno -100 +14900
con prob 99/100 1/100

notero' che $E[X] = E[Y] = 50$, ma $\sigma(X) < \sigma(Y)$: il secondo investi-

mento e' piu' ''rischioso'' del primo.

Se, nella formula della speranza, sostituiamo i valori

$p_i = P[X = x_i]$

Speranza condizionale

con le probabilita' condizionali rispetto ad un evento A, otterremo una stima di X sotto l'ipotesi del verificarsi di A. Chiameremo SPERANZA CONDIZIONALE DI X DATO A la quantita'

$$E[X/A] = \sum_i x_i \; P[(X=x_i)/A].$$

Se $H_1, H_2,.., H_n$ e' una partizione di Ω in eventi misurabili e X e' una variabile aleatoria, e'

$$E[X] = E[X/H_1]P(H_1) + E[X/H_2]P(H_2) + .. + E[X/H_n]P(H_n)$$

formula analoga a quella riportata alla fine del paragrafo 1.2 e altrettanto utile.

1.5 SCHEMA DELLE PROVE RIPETUTE

Schema di Bernoulli

Consideriamo un ''esperimento'' aleatorio (es. lancio di un dado regolare) nell'ambito del quale si possa realizzare un evento S che chiameremo ''Successo'' (es. ''uscita del numero 6''). Sia ''p'' la probabilita' di successo e q = 1 - p la probabilita' di $\mathscr{C}(S)$ = F ovvero la probabilita' di ''Fallimento''.

Se pensiamo di ripetere l'esperimento piu' volte, sempre nelle stesse condizioni, cosi' che l'esito di una prova non venga influenzato da quanto successo nelle precedenti, realizziamo quello che nel calcolo delle probabilita' e' detto SCHEMA DELLE PROVE RIPETUTE (s.p.r.) o SCHEMA DI BERNOULLI di parametro ''p''.

In uno s.p.r. si considera l'indefinita ripetizione dell'esperimento, nel senso che non viene posto, a priori, alcun limite al numero delle prove da eseguire.

Esempi di s.p.r. sono le estrazioni dei numeri del lotto nelle diverse settimane, o i successivi ''giochi'' alla roulette.

Se indichiamo con S_i l'evento {''Successo'' alla i-esima prova}, la successione $S_1,S_2,..,S_n,..$ e' una famiglia (numerabile) completamente indipendente.

PRIMO PRINCIPIO DELLE PROVE RIPETUTE

In uno s.p.r. di parametro $p > 0$, prima o poi si verifichera' un successo.

In effetti, l'evento

$E = \{$sempre e solo fallimenti$\} \subseteq \{$fallimenti nelle prime N prove$\} = E_N$,
quindi

$$P(E) \leqslant P(E_N) = P[\underbrace{F \cap F \cap .. \cap F}_{N \text{ volte}}] = \underbrace{qq..q}_{N \text{ volte}} = q^N.$$

Se $p > 0$ e' $q < 1$ e q^N puo' essere reso arbitrariamente piccolo ($\forall \epsilon > 0 \ \exists N: q^N < \epsilon$) per cui
$P(E) = 0$ e
$P(\mathscr{C}(E)) = P[$prima o poi si verifichera' un successo$] = 1$.

Il risultato non dipende dal particolare valore di ''p'', purche' positivo, ed e' possibile concepire delle situazioni che hanno dell'incredibile:

Paradosso di Borel

PARADOSSO DI BOREL: una scimmia che batta a caso i tasti di una telescrivente, scrivera', prima o poi, il testo esatto delle Divina Commedia. (Si considera come prova la battitura di un numero di caratteri pari a quelli della Divina Commedia -compresi spazi, interlinee ecc.- e come ''Successo'' la redazione esatta del testo. La probabilita' di successo in ogni singola prova e' assai ridotta, ma non nulla).

Certi di ottenere un successo, possiamo chiederci quanto dovremo attenderlo. Indichiamo con T il numero d'ordine della prima prova in cui si realizza l'evento S; T e' una variabile aleatoria:

T vale	se lo schema inizia con
1	S
2	FS
3	FFS
4	FFFS
...	...

Di questa variabile aleatoria avra' senso calcolare il valore atteso $E[T]$.
Posto $H = \{$Successo alla 1' prova$\}$
$K = \{$Fallimento alla 1' prova$\}$
(H,K) e' una partizione di Ω, per cui (vedi paragrafo 1.4)

$$E[T] = E[T/H] \cdot P(H) + E[T/K] \cdot P(K).$$

Ma,
$E[T/H] = 1$: nell'ipotesi H e' $T = 1$
$E[T/K] = 1 + E[T]$: nell'ipotesi K, la prima prova viene ''spre-

cata'', dalla seconda bisognera' attendere mediamente altre E[T] prove;

$P(H) = p \; ; \; P(K) = q$

quindi:

$$E[T] = 1 \cdot p + (1 + E[T]) \cdot q$$

da cui, ricordando che $p+q=1$, si ricava:

$$E[T] = 1/p.$$

SECONDO PRINCIPIO DELLE PROVE RIPETUTE

In uno s.p.r. di parametro $p>0$, il numero medio delle prove da eseguire per avere il primo successo e' $1/p$.

Ad esempio, per ottenere il primo 6 nel lancio del dado, occorrera' lanciarlo, mediamente, 6 volte.

Il paradosso di Borel diviene ora piu' accettabile: la scimmia scrivera' la Divina Commedia, ma p e' talmente piccola che il tempo medio di attesa e' astronomico.

Se, di uno s.p.r., si eseguono esattamente ''n'' prove, ci si puo' chiedere quale sia la probabilita' di avere un numero prefissato ''k'' ($0 \leqslant k \leqslant n$) di successi. Poiche' il calcolo di tali probabilita' richiede alcune nozioni di calcolo combinatorio, a questa domanda si rispondera' nel 4' capitolo dove si affrontano appunto argomenti di calcolo combinatorio e la distribuzione binomiale (si veda anche il problema ''Compito in classe'' nel 2' capitolo).

1.6 LEGGE DEI GRANDI NUMERI

In questo paragrafo si affrontano tre argomenti:

(a) Disuguaglianza di Cebiscev
(b) Calcolo di media e varianza della frequenza di successi in uno s.p.r.
(c) Legge dei grandi numeri (in relazione ad uno s.p.r.)

Le dimostrazioni relative ai punti (a) e (b) sono piuttosto difficili e possono essere ignorate senza che cio' pregiudichi la comprensione del discorso.

Sia X una variabile aleatoria di legge (x_i, p_i), di media $\mu = E[X]$ e varianza $\sigma^2 (X)$.

Se λ e' un numero reale positivo arbitrario, vale la seguente disuguaglianza (DISUGUAGLIANZA DI CEBISCEV):

Disuguaglianza di Cebiscev

$$P[|X - \mu| \geqslant \lambda] \leqslant \frac{\sigma^2(X)}{\lambda^2}$$

In termini discorsivi, la probabilita' che la variabile aleatoria assuma valori che si discostano dalla media ''piu' di λ'' non supera $\sigma^2(X)/\lambda^2$.

Cio' significa che, se $\sigma^2(X)$ e' ''piccola'', e' poco probabile che X si discosti ''molto'' da μ. Se, ad esempio, poniamo $\mu = 100$ e $\sigma^2(X)=1$, per $\lambda=10$ si ha:

$$P[|X - 100| \geqslant 10] \leqslant 1/100 :$$

la probabilita' che X assuma valori inferiori a 90 o superiori a 110 e' minore di 1/100.

DIMOSTRAZIONE

Fissato $\lambda>0$, indichiamo con x'_i i valori x_i per i quali e' $|x'_i - \mu| \geqslant \lambda$; si avra' anche $(x'_i - \mu)^2 \geqslant \lambda^2$.

L'insieme $(|X - \mu| \geqslant \lambda)$ e' l'unione disgiunta di $\{x'_i\}$ quindi $P[|X - \mu| \geqslant \lambda] = \sum_i p'_i$.

Si ha dunque

$$\sigma^2(X) = \sum_i (x_i - \mu)^2 p_i \geqslant \sum_i (x'_i - \mu)^2 p'_i \geqslant \lambda^2 \sum_i p'_i = \lambda^2 \cdot P[|X - \mu| \geqslant \lambda].$$

Dividendo per λ^2 si ottiene la disuguaglianza di Cebiscev.

In uno s.p.r. di parametro ''p'', sia F_n la variabile aleatoria ''frequenza di successo nelle prime n prove'' = Numero di successi/n.

F_n ha media p e varianza pq/n.

DIMOSTRAZIONE

Se indichiamo I_k la variabile aleatoria che vale:
1 se alla prova k si e' avuto successo
0 se alla prova k si e' avuto fallimento

$$\text{sara' } F_n = \frac{1}{n}(I_1 + I_2 + .. + I_n).$$

Avremo:

$E[I_k] = 1 \cdot p + 0 \cdot q = p$ e

$$E[F_n] = \frac{1}{n} (E[I_1] + E[I_2] + .. + E[I_n]) = \frac{n \cdot p}{n} = p$$

ove si e' tenuto conto che E e' un'operatore lineare: $E[aX + bY] = aE[X] + bE[Y]$ con X,Y var. aleatorie e a,b numeri reali.

$$\sigma^2(F_n) = E[(F_n-p)^2] = E[F_n^2 + p^2 - 2F_n \cdot p] =$$
$$= E[F^2_n] + p^2 - 2pE[F_n] = E[F^2_n] + p^2 - 2p^2 = E[F^2_n] - p^2$$

Vediamo come calcolare $E[F^2_n]$:

$$F^2_n = \frac{1}{n^2} (I_1 + I_2 + .. I_n)^2 = \frac{1}{n^2} (I^2_1 + I^2_2 + .. + I^2_n + \sum_{h \neq k} I_h \cdot I_k)$$

ma $I^2_k = I_k$ ($0^2 = 0$ e $1^2 = 1$) quindi $E[I^2_k] = E[I_k] = p$.

Per $h \neq k$, $I_h \cdot I_k$ vale 1 solo se si verifica un successo sia alla prova numero h sia alla prova numero k. Per l'indipendenza, cio' avviene con probabilita' $p \cdot p = p^2$, quindi

$$E[I_h I_k] = 1 \cdot p^2 + 0(1 - p^2) = p^2.$$

Allora

$$E[F^2_n] = \frac{1}{n^2} [E[I^2_1] + E[I^2_2] + .. + E[I^2_n] + \sum_{h \neq k} E[I_h I_k]] =$$
$$= \frac{1}{n^2} [p + p + .. + p + \sum_{h \neq k} p^2]$$

l'ultima somma contiene n(n-1) addendi uguali a p^2 per cui:

$$E[F^2_n] = \frac{1}{n^2} [np + n(n-1)p^2] = \frac{1}{n^2} [np + n^2p^2 - np^2] = \frac{p(1-p)}{n} + p^2.$$

In definitiva:

$$\sigma^2(F_n) = \frac{p(1-p)}{n} + p^2 - p^2 = \frac{pq}{n}$$

Poiche' $\sigma^2(F_n) = pq/n$, la varianza diminuisce al crescere del numero delle prove; applicando la disuguaglianza di Cebiscev a F_n, si ha (LEGGE DEI GRANDI NUMERI):

Legge dei grandi numeri

$$\forall \lambda > 0 \qquad P[|F_n - p| \geqslant \lambda] \leqslant \frac{pq}{\lambda^2 n} :$$

comunque si scelga un numero $\lambda > 0$, la probabilita' che la frequenza di successo si discosti da ''p'' per piu' di λ, puo' essere resa arbitrariamente piccola, pur di considerare un numero di prove abbastanza grande.

Passando al complementare, e' :

$$\forall\lambda > 0 \qquad \underline{P}\,[p-\lambda < F_n < p+\lambda] > 1 - \frac{pq}{\lambda^2 \cdot n}.$$

Cio' significa che, comunque si scelga un intervallo $(p-\lambda, p+\lambda)$, anche molto piccolo, la probabilita' che la frequenza di successo cada in questo intervallo puo' essere resa arbitrariamente prossima ad 1 scegliendo un numero di prove sufficientemente elevato.

Osservazioni

Attraverso una serie di osservazioni, rivediamo i risultati degli ultimi due paragrafi applicandoli al seguente esempio:

Un dado regolare viene lanciato piu' volte; chiamiamo ''Successo'' l'evento

S = {uscita del numero 6}

I successivi lanci forniscono un esempio di s.p.r. di parametro $p = P(S) = 1/6$.

1. Poiche' $p > 0$, se ripetiamo l'esperimento, prima o poi otterremo un 6 (primo principio delle prove ripetute)

2. Quante volte occorre tirare il dado per avere il primo successo? Si potrebbe avere un 6 al primo lancio, oppure dover attendere il secondo o il terzo o il quarto ecc. Il numero di lanci ''dipende dal caso'' ma, per il secondo principio, occorreranno mediamente $1/p = 6$ tiri.

3. Pensiamo di lanciare il dado un numero prefissato ''n'' di volte; contiamo il numero ''k'' di successi ottenuti; sara' $0 \leqslant k \leqslant n$.

Calcoliamo $F_n = k/n$ che chiamiamo ''frequenza di successo'' nelle n prove; avremo $0 \leqslant F_n \leqslant 1$ ($F_n = 0$ se non si e' verificato alcun successo; $F_n = 1$ se in tutti i lanci e' uscito 6). F_n e' una variabile aleatoria; calcoliamone il valore medio e la varianza:

$E[F_n] = p = 1/6$; $\sigma^2(F_n) = pq/n = 5/(36n)$

4. Cosa si puo' dire di F_n quando n diviene molto grande? La legge dei grandi numeri applicata a questo esempio diviene:

$$\forall\lambda > 0 \qquad \underline{P}\left[\frac{1}{6} - \lambda < F_n < \frac{1}{6} + \lambda\right] > 1 - 5/(36\lambda^2 n).$$

Prendiamo $\lambda = 1/100$; e' $1/6 \simeq 0.167$; $1/6 - \lambda \simeq 0.157$; $1/6 + \lambda \simeq 0.177$ e $5/(36\lambda^2) \simeq 1389$
quindi, approssimativamente,

$$P[0.157 < F_n < 0.177] > 1 - 1389/n.$$

Supponiamo di eseguire 100000 prove ($n = 100000$), avremo:

$$P[0.157 < F_{100.000} < 0.177] \geq 0.986.$$

Eseguendo 100000 prove, la probabilita' che la frequenza di successo sia compresa tra 0.157 e 0.177 e' piu' del 98% !

All'aumentare del numero delle prove, la frequenza ''si avvicina'' alla probabilita' nel senso che ''grandi differenze'' tra frequenza e probabilita' divengono sempre meno probabili (attenzione, questo non significa che $P(F_n = p)$ si avvicini a 1; al contrario, la probabilita' che la frequenza sia ESATTAMENTE uguale a ''p'' tende a zero al crescere del numero delle prove).

La stretta parentela tra ''frequenza'' e ''probabilita'' giustifica, almeno in parte, la concezione frequentista di cui si discuteva nel paragrafo 1 anche se, definire la probabilita' come particolare frequenza, crea comunque delle difficolta'. Con le dovute cautele, si potra' utilizzare la frequenza come stima di una probabilita' incognita anche perche', quando si ricorre a probabilita' di questo tipo, piu' che il valore esatto interessa ''non sbagliarsi di molto''.

Guadagno medio teorico

Se, in uno s.p.r., all'evento ''Successo'' e' associato un guadagno ''x'', eseguendo ''n'' prove ed ottenendo ''k'' successi, si guadagnera' complessivamente $G = kx$. Il guadagno medio nella singola prova e' dunque $GM = kx/n = F_n x$.

Sostituendo F_n con p, si ottiene $px = E[X]$ dove X e' la variabile aleatoria di legge

x_i : x 0

p_i : p 1-p .

Si spiega ora l'interpretazione di E[X] come ''guadagno medio teorico'' nei termini descritti nel paragrafo 1.4.

1.7 SIMULAZIONI

La maggior parte dei programmi associati agli esercizi contenuti nel prossimo capitolo realizzano delle simulazioni.

Ciascuno dei problemi proposti si risolve calcolando certe quantita'; generalmente si tratta delle probabilita' di particolari eventi e/o dei valori medi di certe variabili aleatorie.

Dopo aver risolto l'esercizio ''per via teorica'', perche' non mettere alla prova le nostre conclusioni? Probabilita' e valori attesi possono essere approssimati dalle frequenze in un opportuno s.p.r.; purche' si eseguano molte prove in condizioni, per quanto possibile, immutate.

La ripetizione di un esperimento aleatorio (es. lancio di un dado) per centinaia o migliaia di volte e' un lavoro lungo e noioso, se eseguito manualmente; un invito a nozze per l'instancabile calcolatore. Quasi tutti i calcolatori dispongono di un generatore di numeri pseudocasuali (vedi paragrafo 1.8) che puo' essere utilizzato, con le opportune manipolazioni, per simulare l'esito di una prova di uno s.p.r.

Un programma di simulazione sara' dunque un programma nel quale viene ripetuto l'esperimento aleatorio (simulato) e nel quale vengono calcolate certe frequenze da confrontarsi con i valori attesi.

Quando il problema riguarda un gioco in cui chi partecipa puo' scegliere tra differenti strategie, e' possibile simulare contemporaneamente il comportamento dei diversi giocatori. Se abbiamo calcolato che la strategia migliore e' quella adottata dal sig. Rossi, ci fara' piacere vedere che, di fatto, il nostro eroe si piazza in testa alla classifica, superando, magari di misura, avversari che adottavano strategie molto promettenti.

Naturalmente non ci aspetteremo una perfetta sovrapposizione tra risultati teorici e valori simulati, ma quella ''giusta approssimazione'' tra frequenza e probabilita' nei termini previsti dalla legge dei grandi numeri (cosi', se ci aspettiamo che A vinca piu' spesso di B e succede il contrario, non e' il caso di buttare il dischetto dalla finestra).

Vediamo, con un semplicissimo esempio, come si sviluppa il discorso problema - soluzione teorica - simulazione.

Problema

Flavio e Daniela, due bambini di sette anni, decidono di ripassare le tabelline giocando con due dadi, uno rosso e uno bianco. Flavio dice a Daniela: ''lanciamo i dadi e moltiplichiamo i due numeri; se il risultato e' pari tu dai una caramella a me, se e' dispari la do io a te''.

Fa bene Daniela ad accettare la proposta di Flavio ?

Soluzione

Daniela vince se il prodotto dei valori e' dispari e cio' avviene solo se su entrambi i dadi esce un numero dispari; se

DR = {uscita dispari sul dado rosso} e
DB = {uscita dispari sul dado bianco} e'
P(DR) = P(DB) = 1/2.

I due lanci sono indipendenti per cui

P[Daniela vince] = P (DR $\cap$ DB) = P(DR)P(DB) = (1/2)(1/2) = 1/4.

Quando Daniela perde, Flavio vince, quindi

P[Flavio vince] = 1-1/4 = 3/4

(oppure, dei 4 casi ugualmente possibili: PP, PD, DP, DD, i primi 3 sono favorevoli a Flavio e solo l'ultimo a Daniela).

Daniela farebbe bene a rifiutare la proposta di Flavio.

Simulazione

Cosa succede se i due bambini si mettono a giocare?

Un programma di simulazione andra' cosi' strutturato:

1- Richiesta del numero N di partite che si vogliono simulare
2- Ciclo principale: per ogni partita, cioe' da 1 a N,
   2.1 generazione casuale (o, meglio, pseudocasuale) del valore DR, risultato del lancio del dado rosso (DR sara' un numero da 1 a 6 costruito in modo che ogni valore abbia la stessa probabilita' di essere scelto)
   2.2 operazione analoga per il dado bianco, valore DB
   2.3 calcolo del prodotto PR = DR*DB
   2.4 se PR e' pari, viene incrementato di una unita' il contatore VF, contatore delle partite vinte da Flavio
   2.5 se PR e' dispari, viene incrementato di una unita' il contatore VD (partite vinte da Daniela)
   2.6 se le partite simulate sono meno di N, si ritorna al punto 2.1
3- A questo punto VF e VD contengono il totale delle partite vinte da Flavio e Daniela rispettivamente; le frequenze saranno VF/N e VD/N, valori da confrontare con quelli attesi: 0.75 e 0.25

Per chi ha un minimo di familiarita' con il Basic, non dovrebbe essere difficile realizzare in pratica quanto descritto; la redazione del programma viene lasciata come esercizio.

1.8 NUMERI PSEUDOCASUALI

L'esigenza di disporre di lunghe sequenze di numeri casuali ha spinto i ricercatori a studiare una serie di procedure per la loro generazione. La maggior parte dei generatori lavora, grosso modo, secondo il seguente algoritmo:

1- viene richiesto un numero x (seme) appartenente ad un certo

intervallo (a,b)
2- viene calcolato un numero $y = f(x)$ dove ''f'' e' costruita in modo che $y \epsilon (a,b)$
3- il valore y viene sostituito a x
4- si rientra al punto 2 per il calcolo di un nuovo numero y.

I numeri cosi' generati vengono detti ''pseudocasuali'', in quanto, non solo la sequenza si ripete identica dopo un certo numero di passi (PERIODO), ma ogni valore risulta perfettamente determinato dal precedente.

Praticamente pero', quello che interessa e' che la successione ''si comporti'' come un'autentica successione casuale; il generatore viene ritenuto soddisfacente se la sequenza che produce supera un certo numero di test di casualita' (test che ne misurano appunto il ''grado di casualita''').

Test di casualita'

Supponiamo di voler simulare i lanci di un dado regolare e di dover scegliere una delle seguenti successioni:

Sequenza 1

1 2 3 4 5 6 1 2 3 4
5 6 1 2 3 4 5 6 1 2
3 4 5 6 1 2 3 4 5 6

Sequenza 2

1 4 1 5 2 6 5 3 5 3
2 3 4 6 2 6 4 3 3 3
2 5 2 4 1 1 6 3 3 5

Mentre la prima sequenza appare a prima vista insoddisfacente (perche', ad esempio, $x_{i+1} \neq x_i \forall i$), la seconda sembra del tutto casuale, nonostante sia semplicemente costituita dalle prime 30 cifre dello sviluppo decimale di π (dove sono state ignorate quelle non comprese tra 1 e 6).

Una trattazione dei test statistici di casualita' richiede conoscenze matematiche superiori ed esula dagli scopi di questo libro; il lettore interessato potra' consultare [7] e [9] dove sono descritti i test piu' noti (test di uniformita', test seriale, test del gap, test del poker, test della distanza).

Congruenza lineare

Per quanto riguarda la generazione di numeri pseudocasuali uniformemente distribuiti (ovvero scelti indifferentemente ''a caso'' in un intervallo), il metodo piu' usato e' quello della CONGRUENZA LINEARE dove la relazione ricorrente e' data da:

$$X_{n+1} = (A \cdot X_n + C) \mod M$$

(X_{n+1} e' uguale al resto della divisione tra $A \cdot X_n + C$ e M)

in cui A,C,M sono numeri interi e X_0 e' compreso tra 0 e M. Per ottenere una buona successione con periodo sufficientemente esteso, A, C ed M vanno scelti in modo opportuno (si veda [7], pag.60), in particolare M deve essere piuttosto elevato, dato che il periodo non puo' superare M. Ecco un esempio:

```
1 C=99991:A=24298:M=199017:INPUT X
2 N=A*X+C:X=N-INT(N/M)*M:PRINT X:GO TO 2
```

Dividendo per M i valori cosi' generati, si otterranno numeri pseudocasuali dell'intervallo [0,1).

La funzione RND in Basic

In Basic, la funzione RND fornisce appunto un valore pseudocasuale compreso tra 0 e 1; generalmente X_0 non viene richiesto dall'esterno, ma viene automaticamente assegnato utilizzando valori forniti dall'orologio interno del calcolatore. La sintassi dell'istruzione prevede anche la possibilita' di ottenere la stessa sequenza in momenti diversi (basta iniziare con lo stesso seme); un'opzione che consente di ripetere in modo identico una simulazione (e, in effetti, la RIPRODUCIBILITA' e' un requisito importante di un buon generatore di numeri pseudocasuali).

CAPITOLO 2

PROBLEMI DI CALCOLO DELLE PROBABILITÀ

In questo capitolo sono raccolti 17 problemi; a ciascun problema e' collegato un programma (a volte piu' di uno) che calcola certi valori o realizza delle simulazioni, come specificato alla fine del capitolo precedente.

Ho cercato, per quanto possibile e a costo di qualche ripetizione, di mantenere gli esercizi ''indipendenti'' l'uno dall'altro: ogni problema e' un paragrafo a se' e contiene i richiami agli argomenti teorici necessari alla comprensione della soluzione (e, dove serve, qualcosa di piu').

In questo modo il lettore puo' scegliere liberamente gli esercizi che ritiene piu' interessanti; giovera' sapere che sono disposti in ordine crescente di difficolta' (i numeri 13, 14 e 15 sono piuttosto difficili).

A volte il problema consiste nel determinare la miglior strategia per un certo gioco. In questi casi, e' bene soffermarsi sul testo e procedere alla lettura della soluzione solo dopo aver ben compreso il meccanismo del gioco; provate sempre a formulare una risposta che, intuitivamente, vi sembra plausibile e, perche' no, a giocare qualche partita.

Gli ultimi due esercizi (''Il linguaggio'' e ''L'opinione pubblica'') non sono dei problemi veri e propri, ma, per cosi' dire, delle curiosita': un microesempio di come modelli probabilistici possano essere utilizzati (e di fatto lo sono) in campi davvero insospettati.

Per quanto riguarda i programmi di simulazione, va segnalato che il tempo di elaborazione puo' crescere considerevolmente all'aumentare del numero delle prove: nei primi esperimenti, converra' non esagerare nelle richieste.

Con pochissime eccezioni, i programmi registrati ricalcano quelli inseriti nel testo (la velocita' e' stata leggermente sacrificata a favore della ''leggibilita''' dei listati); una routine sonora si incarica

di segnalare la conclusione dell'elaborazione.

2.1 CARLO E ANDREA

Un programma associato:

"CARLO E ANDREA"

Un dado regolare a sei facce e' stato modificato: su due facce e' rappresentata la cifra 1, su altre due la cifra 2 e sulle rimanenti due la cifra 3.

Carlo e Andrea si alternano nel lancio del dado: il primo che non ottiene un valore maggiore di quello tirato dall'avversario nel turno precedente, perde il gioco (e l'altro vince). Inizia Carlo.

Qual e' la probabilita' di vittoria di Carlo? Quale quella di Andrea? Qual e' il numero medio di lanci per ogni singola partita? (durata media del gioco)?

ESEMPI

(C = tiro di Carlo; A = tiro di Andrea; V = vincitore C/A; NL = numero lanci)

1. C = 1; A = 1 V = C; NL = 2
2. C = 3 V = C; NL = 1
3. C = 1; A = 3 V = A; NL = 2
4. C = 1; A = 2; C = 3 V = C; NL = 3

SOLUZIONE

Il problema, proprio perche' semplicissimo, suggerisce immediatamente queste considerazioni:

- Il gioco non puo'terminare in parita'(si gioca "al rialzo")
- Carlo e' decisamente avvantaggiato
- Il numero complessivo di lanci necessari a "chiudere" una partita va da 1 (esempio 2) a 3 (esempio 4).

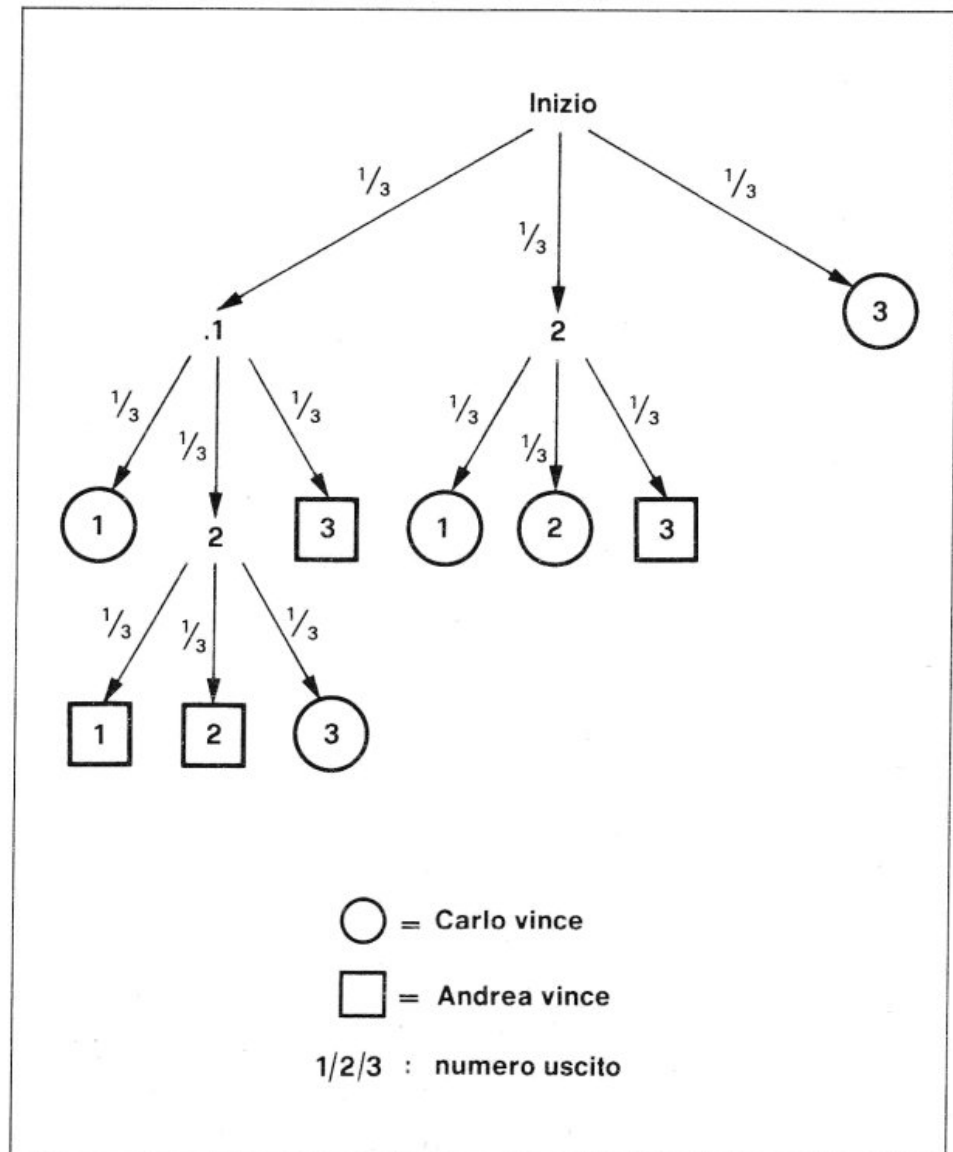
Il limitato numero di casi consente un elenco di tutte le partite possibili: tre lanci del "nostro" dado originano 27 partite; a fianco di ciascuna e' scritta l'iniziale del vincitore e il numero di lanci effettuati per concludere la gara. Ad esempio, "132 A 2" significa che Carlo ottiene 1 al 1' lancio, Andrea 3 al 2' lancio; la partita si conclude quindi dopo 2 lanci con la vittoria di Andrea (A).

111 C 2 211 C 2 311 C 1
112 C 2 212 C 2 312 C 1

113 C 2	213 C 2	313 C 1
121 A 3	221 C 2	321 C 1
122 A 3	222 C 2	322 C 1
123 C 3	223 C 2	323 C 1
131 A 2	231 A 2	331 C 1
132 A 2	232 A 2	332 C 1
133 A 2	233 A 2	333 C 1

Le ''situazioni favorevoli'' a Carlo sono dunque 19, quelle favorevoli ad Andrea 8. Le probabilita' richieste sono:
p = P[Carlo vince il gioco] = 19/27 ($\simeq$ 0.704)
q = P[Andrea vince il gioco] = 8/27 ($\simeq$ 0.296)

La ''durata media del gioco'' (intesa come numero di lanci) e' data dalla media dei valori della terza colonna ovvero 48/27 = 16/9 $\simeq$ 1.778.

Fig. 4 *Grafo per la soluzione del problema 2.1*

Un approccio di questo tipo e' attuabile solo se il numero di ''casi possibili'' si mantiene moderato; alla soluzione si giunge molto piu' rapidamente utilizzando il grafo di fig.4 (vedi paragrafo 1.2), ottenendo:

P[vince Carlo] = 1/3 + 3/9 + 1/27 = 19/27
P[vince Andrea] = 2/9 + 2/27 = 8/27.

Un discorso analogo puo' essere fatto per il numero medio di lanci (paragrafo 1.4):

1 lancio con probabilita' 1/3
2 lanci con probabilita' 5/9
3 lanci con probabilita' 1/9

da cui:

1 (1/3) + 2 (5/9) + 3 (1/9) = 16/9 .

SIMULAZIONE

I risultati precedenti suggeriscono che, su 100 partite, Carlo totalizzera' circa 70 vittorie contro le 30 di Andrea; il dado verra' lanciato mediamente 1.78 volte per concludere ogni singola gara.

Nel programma che segue, RND(1) fornisce un numero casuale tra 0 e 1 e la funzione introdotta alla linea 10 determina un intero pseudo-casuale tra 1 e 3. L'utilizzo di variabili (U, D, T) in luogo di costanti (1, 2, 3) riduce il tempo di elaborazione nel ciclo principale (30-90).

Il programma richiede il numero di partite da giocare e fornisce la frequenza di successo di Carlo (che ''tendera''' a 0.704), quella di Andrea e il numero medio di lanci per partita.

Le variabili utilizzate sono :
NP = numero di partite che si intendono simulare
L1, L2, L3 = risultato del 1', 2' e 3' lancio
C = contatore vittorie di Carlo
A = contatore vittorie di Andrea
L = contatore numero totale dei lanci eseguiti.

Esempi :

QUANTE PARTITE? 500
FREQ. DI SUCCESSO PER CARLO: 0.704
FREQ. DI SUCCESSO PER ANDREA: 0.296
NUMERO MEDIO DI LANCI PER PARTITA: 1.782

QUANTE PARTITE? 2000
FREQ. DI SUCCESSO PER CARLO: 0.7045
FREQ. DI SUCCESSO PER ANDREA: 0.2955
NUMERO MEDIO DI LANCI PER PARTITA: 1.772

Listato

```
  5 U=1:D=2:T=3
 10 DEF FND(X)=INT(RND(U)*T)+U
 20 INPUT''QUANTE PARTITE'';NP
 30 FOR I=U TO NP
 40 L1=FND(U):IF L1=T THEN C=C+U:L=L+U:GO TO 90
 50 L2=FND(U):IF L2<=L1 THEN C=C+U:L=L+D:GO TO 90
 60 IF L2=T THEN A=A+U:L=L+D:GO TO 90
 70 L=L+T:L3=FND(U):IF L3=T THEN C=C+U:GO TO 90
 80 A=A+U
 90 NEXT I
100 PRINT''FREQ. CARLO'';C/NP
110 PRINT''FREQ. ANDREA'';A/NP
120 PRINT''NUM. MEDIO LANCI PER PARTITA'';L/NP
```

2.2 IL COLLEZIONISTA

Un programma associato:

''COLLEZIONISTA''

Ogni scatola di detersivo contiene una figurina da incollare su un album in cui trovano posto N differenti figurine.

Quante scatole di detersivo si dovranno comperare mediamente per completare la raccolta? Qual e' la probabilita' di cavarsela con esattamente N acquisti?

Si tratta di un problema classico (si veda [15], pag. 72), che offre lo spunto per due programmi: un breve programma per il calcolo della soluzione e un programma di simulazione.

Si ricorda che in uno schema di Bernoulli (s.p.r.) di parametro $p>0$, il ''tempo medio di attesa'' per il primo successo e' $1/p$.

SOLUZIONE

La prima scatola va sempre bene, dato che la figurina trovera' posto sull'album. Il secondo acquisto e' la prima prova di uno schema delle prove ripetute in cui ''successo'' e' trovare una figurina diversa dalla prima. La probabilita' di successo e' $p=(N-1)/N$ e

il numero medio di scatole da acquistare e' allora N/(N-1).

In modo analogo si procede per gli ulteriori acquisti:

FIGURINA UTILE	CASI FAVOREVOLI	PROB. DI SUCCESSO	NUM. MEDIO DI SCATOLE DA ACQUISTARE
1'	tutti	1	1
2'	N - 1	(N - 1)/N	N/(N - 1)
3'	N - 2	(N - 2)/N	N/(N - 2)
...	...	...	...
N-esima	1	1/N	N

Il numero medio di scatole da acquistare e':

$$M = 1 + \frac{N}{N-1} + \frac{N}{N-2} + .. + \frac{N}{1} = N(1 + \frac{1}{2} + \frac{1}{3} + .. + \frac{1}{N}).$$

Completare la raccolta con esattamente N acquisti significa collezionare N successi nelle prime N prove. La probabilita' di tale evento e' data da:

$$P = 1 \cdot \frac{N-1}{N} \cdot \frac{N-2}{N} \cdot \; ... \; \cdot \frac{1}{N} = \frac{N!}{N^N}$$

quantita' questa che, anche per modesti valori di N, e' assai piccola (per esempio, per $N=10$ e' $P=0.000359$).

Il programma 1 calcola il numero medio di acquisti necessari per il completamento di un album di N figurine.

Listato programma 1

```
10 INPUT''QUANTE FIGURINE'';N
20 FOR I=1 TO N:S=S+1/I:NEXT I
30 PRINT''NUMERO MEDIO DI ACQUISTI'';S*N
```

Ecco alcuni valori ottenuti con il programma 1:

N	M circa
10	29
(*) 52	236
100	519
200	1176
1000	7485

(*) Molti anni fa erano in vendita scatole di formaggini con una carta da gioco di un mazzo di 52 carte in ogni confezione

Simulazione

Nel programma 2 si realizza la simulazione. All'istruzione 5 viene richiesto da quante figurine e' composto l'album (NF). Il vettore A(..), dimensionato alla linea 10, rappresenta l'album: se A(I) = 0 la casella I e' vuota, se A(I) = 1 la casella I e' riempita dalla relativa figurina.

Sempre nella 10, viene richiesto il numero di simulazioni, ovvero quanti collezionisti partecipano alla raccolta.

Per ogni collezionista (ciclo 30-90), la variabile A contiene, in ogni momento, il numero di acquisti effettuati; la variabile N il numero di figurine presenti sull'album. Inizialmente (istruzioni 35 e 36) questi valori vengono azzerati e l'album ''svuotato''.

All'istruzione 40, F contiene il numero della figurina estratta (al solito RND(1) fornisce un valore pseudo-casuale tra 0 e 1) e, alla 50, viene incrementato il numero degli acquisti. Se la figurina e' utilizzabile, viene incrementato N e la figurina viene inserita nell'album (istruzione 60).

L'istruzione 70 si occupa di controllare se la raccolta e' completa: in caso negativo si ritorna alla linea 40 per un nuovo acquisto, in caso affermativo viene stampato il numero del collezionista e il numero di acquisti sostenuti.

La variabile T dell'istruzione 90 contiene il totale di tutti gli acquisti, che viene aggiornato.

All'uscita del ciclo 30-90, viene stampato il numero medio di acquisti effettuati da ogni collezionista. Tale valore sara', per un numero di simulazioni sufficientemente elevato, abbastanza prossimo a quello previsto dal programma 1.

Listato programma 2

```
  5 INPUT''NUMERO FIGURINE'';NF
 10 DIM A(NF):Z=0:U=1:INPUT S
 30 FOR K=U TO S
 35 FOR L=U TO NF:A(L)=Z:NEXT L
 36 A=Z:N=Z
 40 F=INT(RND(U)*NF)+U
 50 A=A+U
 60 IF A(F)=Z THEN N=N+U:A(F)=U
 70 IF N=NF THEN 90
 80 GO TO 40
 90 PRINT K,A:T=T+A:NEXT K
120 PRINT''VALORE MEDIO'';T/S
```

Esempio di output del programma 2

NUMERO FIGURINE : 200
NUMERO SIMULAZIONI : 10

COLLEZIONISTA N.	ACQUISTI
1	1059
2	1602
3	1016
4	947
5	1360
6	1178
7	1246
8	904
9	1085
10	1205

VALORE MEDIO : 1166.2

(Il valore teorico per un album di 200 figurine e' 1176).

Nel programma ''COLLEZIONISTA'', registrato sul supporto magnetico, sono condensati i due programmi: inizialmente viene chiesto il numero delle figurine; il programma stampa il numero medio di acquisti e richiede il numero di simulazioni procedendo quindi come nel programma 2.

2.3 IL GIOCO DEL MINI-LOTTO

Un programma associato:

''MINI-LOTTO''

Immagino che quanto segue risultera' ovvio per la maggior parte di coloro che leggono queste pagine, ma la pubblicazione sui giornali di quei numeri del lotto che accumulano molte settimane di ''ritardo'' mi fa pensare che qualche considerazione possa risultare utile.

Le conclusioni a cui giungeremo tramite il modello semplificato

rimangono valide per il normale gioco del lotto (liberandoci dalle complicazioni di carattere combinatorio).

Ecco la favoletta:

Molti cittadini di Bingotown giocano ogni settimana al Mini-lotto. Dal lunedi' al venerdi' e' possibile puntare su uno dei dieci numeri 0, 1, 2,..., 9; al sabato viene eseguita un'estrazione e coloro che hanno giocato il numero sorteggiato vincono. La somma per una giocata e' fissa e piuttosto elevata; normalmente una persona non gioca piu' di un numero a settimana.

Il sig. Rossi e il sig. Bianchi sono fra i piu' incalliti giocatori: ogni settimana si recano al botteghino.

Rossi tiene aggiornata una ''tabella dei ritardi'' Ri: ad ogni numero ''i'' (i=0,1,..,9) associa il numero di settimane trascorse dall'ultima uscita (Es. Se le uscite sono state 3/0/4/5/1/0/9, sara' R0=1, R1=2, R2=7, R3=6, ..., R9=0). Ogni venerdi' Rossi gioca il numero con maggior ritardo (in caso di parita', sceglie il minore; nell'esempio precedente giochera' il 2 piuttosto del 6 o del 7 o 8).

Il sig. Bianchi agisce diversamente: ogni volta, con l'aiuto di un mazzo di carte ben mescolato, sceglie un numero a caso e lo gioca.

Interminabili discussioni nascono tra i due: Rossi sostiene che la propria strategia e' decisamente migliore e accusa Bianchi di ''non usare il cervello''; Bianchi ribatte che il suo metodo e' altrettanto valido e che il ''ritardato'' non e' il numero, ma una persona di sua conoscenza...

Il programma MINI-LOTTO simula il comportamento dei due:
Istr. 5: Inizializzazione; richiesta numero simulazioni (NS)
Istr. 30-90: Loop principale

- 30-40: in base al valore di massimo ritardo R(K), Rossi decide il numero X da giocare (inizialmente e' R(K)=0 per ogni K).
- 40: Y e' il numero che Bianchi intende giocare; E e' il numero estratto.
- 50-60: TR e TB (Totale vincite Rossi/Bianchi) vengono aggiornati in caso di viincita.
- 70-90: Aggiornamento della ''tabella dei ritardi'' (ad uso di Rossi).

Istr. 100-130: Stampa risultati (numero di giocate vinte da Rossi/Bianchi e relative percentuali)

Listato del programma MINI-LOTTO

```
5 Z=0:U=1:N=9:D=10:DIM R(N):INPUT NS
```

```
 30 FOR I=U TO NS:M=R(Z):X=Z:FOR K=
    U TO N:IF R(K)>M THEN M=R(K):X=K
 40 NEXT K:Y=INT(RND(U)*D):E=INT(RND(U)*D)
 50 IF E=X THEN TR=TR+U
 60 IF E=Y THEN TB=TB+U
 70 FOR L=Z TO N:IF L=E THEN R(L)=Z:GO TO 90
 80 R(L)=R(L)+U
 90 NEXT L:NEXT I
100 PRINT'' V.ROSSI '';TR
110 PRINT'' % ROSSI '';100*TR/NS
120 PRINT'' V.BIANCHI'';TB
130 PRINT'' % BIANCHI'';100*TB/NS
```

Esempi di output

QUANTE SIMUL.?500

V. ROSSI 46
% ROSSI 9.2

V. BIANCHI 55
% BIANCHI 11

QUANTE SIMUL.? 1000

V. ROSSI 119
% ROSSI 11.9

V. BIANCHI 100
% BIANCHI 10

Morale della favola: i risultati forniti dalla simulazione confermano che non esiste differenza fra le due strategie; sia Rossi che Bianchi vincono, mediamente, una volta ogni dieci settimane.

A chi vuol saperne di piu', consiglio la lettura dell'articolo ''Una tassa sulle illusioni'' di Vinicio Villani, apparso sul numero di ottobre 1985 della rivista ''Archimede''.

2.4 TIRO AL PIATTELLO

Un programma associato:

''PIATTELLO''

Nelle gare di tiro al piattello, Rambo e' un vero fuoriclasse: in ogni prova colpisce il bersaglio con probabilita' $p=0.97$.

Prima di iniziare, Rambo deve dichiarare quante prove intende eseguire: se non commettera' neppure un errore, vincera' un numero di dollari pari al numero delle prove dichiarate (nulla in caso contrario).

Qual e' la dichiarazione piu' conveniente?

[si suppongono indipendenti gli esiti delle varie prove]

SOLUZIONE

Se ''x'' e' il numero delle prove dichiarate e ''p'' la probabilita' di successo nella singola prova, Rambo guadagna:

''x'' dollari con probabilita' p^x
nulla con probabilita' $1-p^x$.

Fissato $p\epsilon(0,1)$ - i casi $p=0$ e $p=1$ sono di ovvia soluzione -, il guadagno medio e' $G(x)=x\cdot p^x$ e la dichiarazione ideale e' quell'intero x' che rende massima G(x). Consideriamo la differenza

$$\begin{aligned} G(x+1)-G(x) &= (x+1)p^{x+1} - xp^x = \\ &= (x+1)p^{x+1} - xp^{x+1} + xp^{x+1} - xp^x = \\ &= p^{x+1} - xp^x(1-p) = p^x(p - xq) \text{ dove } q=1-p. \end{aligned}$$

Il valore da dichiarare sara' il piu' piccolo intero per cui $G(x+1)\leqslant G(x)$ (dichiarare un solo tiro in piu' peggiora le cose), dunque il primo intero per il quale $p-xq\leqslant 0$: $x'=-INT(-p/q)$ (la funzione INT(X) fornisce il piu' grande intero minore o uguale a X; INT(-12.2)=-13).

In alternativa, considerando x reale, si ha $G'(x)=p^x(1+x\cdot\log p)$ per cui G(x) ammette massimo per $x=1/|\log p|$. Si trattera' allora di scegliere l'intero migliore tra INT(x) e INT(x)+1.

Il programma PIATTELLO determina, in funzione di P, la dichiarazione ideale X ed il relativo guadagno G(X); si passa quindi alla simulazione.

Simulazione

In ogni gara (il loro numero e' NS e N e' il relativo contatore) vengono eseguite X prove (M contatore): La variabile S, numero di successi, viene incrementata solo quando le X prove hanno tutte fornito esito positivo.

Il rapporto X*S/NS esprime il guadagno medio realizzato da Rambo nella singola gara. Per P=0.97, la dichiarazione ideale e' 33 associata ad un guadagno medio teorico di 12.0776 dollari (confronta fig.5).

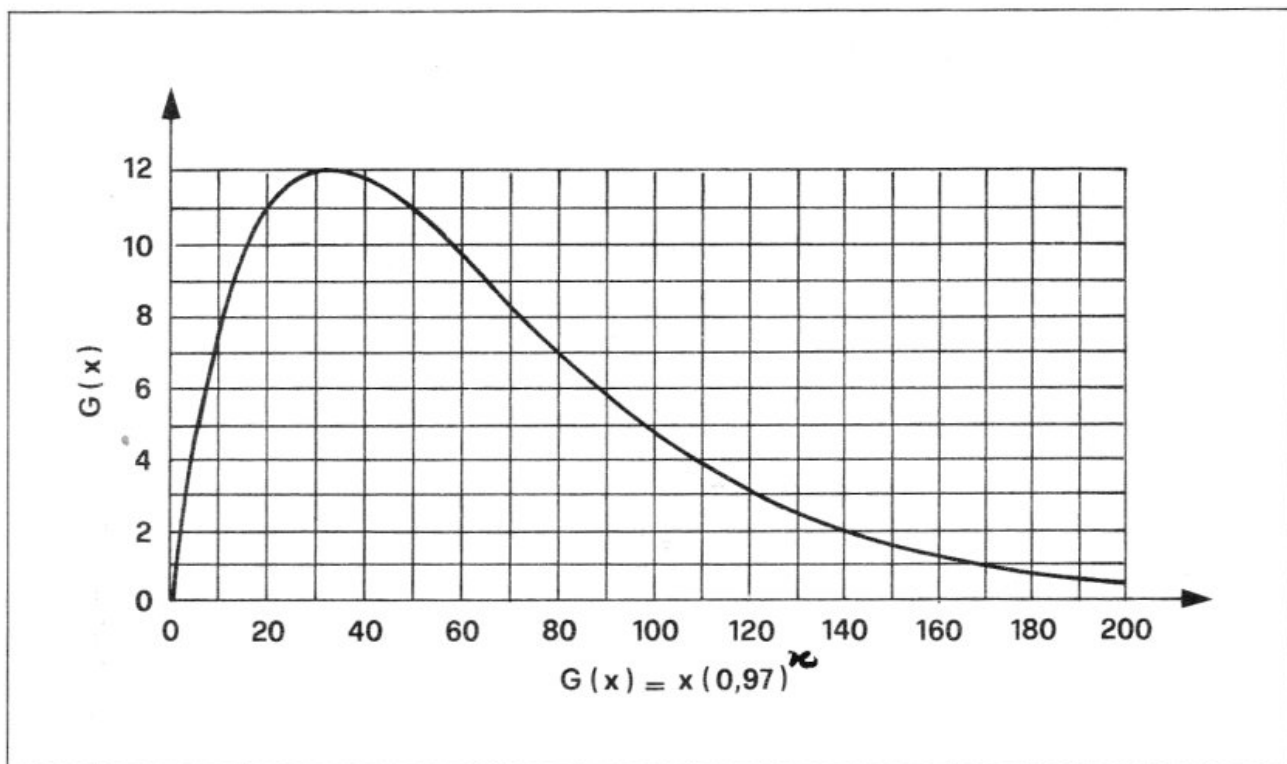



Fig. 5 *Andamento del guadagno in funzione dei tiri dichiarati (problema 2.4)*

Listato del programma PIATTELLO

```
 15 INPUT''PROB. SUCCESSO '';P:X=-INT(-P/(1-P))
 60 PRINT''CONVIENE DICHIARARE'';X:PRINT''GUADAGNO
    MEDIO'';X*P↑X
 70 INPUT NS:FOR N=1 TO NS:M=X
100 IF M THEN IF RND(1)<P THEN M=M-1:GO TO 100
110 IF M=0 THEN S=S+1
120 NEXT N
140 PRINT''GUAD. MEDIO SIMULATO'';X*S/NS
```

Esempi di output del programma PIATTELLO

PROB. SUCCESSO? 0.97
CONVIENE DICHIARARE 33
GUADAGNO MEDIO PREVISTO 12.07761
NUM. SIMULAZIONI? 1000
GUAD. MEDIO OTTENUTO 12.606

PROB. SUCCESSO? 0.5
CONVIENE DICHIARARE 1
GUADAGNO MEDIOPREVISTO 0.5
NUM. SIMULAZIONI? 1000
GUAD. MEDIO OTTENUTO 0.5

2.5 COMPITO IN CLASSE

Due programmi associati:

''BINOMIALE''
''TEST''

Una prova scritta di matematica e' costituita da 10 quesiti ad ognuno dei quali e' associato un gruppo di quattro possibili risposte di cui una esatta e tre errate.

Al candidato e' richiesto di segnalare qual e' l'alternativa che ritiene corretta; l'insegnante valutera' il compito assegnando un punto per ogni risposta esatta.

Augusto, del tutto impreparato sugli argomenti oggetto della prova, sceglie a caso le risposte (in ogni quesito, ciascuna alternativa ha la stessa probabilita' di essere scelta indipendentemente dalla domanda proposta e dalle risposte fornite alle domande precedenti).

Detto X il numero di risposte esatte del compito di Augusto, calcolare:

1. $P[X=k]$ per $k=0,1,2,\ldots,10$
2. $P[X \geqslant 6]$ probabilita' che, nonostante tutto, la preparazione di Augusto venga considerata ''non insufficiente''.

SOLUZIONE

Sia Ai = {Augusto risponde correttamente alla i-esima domanda} $i=1,2,\ldots,10$; e' $p=P(Ai)=1/4$ per ogni i.

Le risposte fornite da Augusto realizzano uno schema delle prove ripetute di parametro $p=1/4$; il problema richiede di calcolare la probabilita' di ottenere esattamente k successi ($k=0,1,\ldots,10$) nelle $n=10$ prove eseguite. Tali probabilita' sono (vedi 4.5 DISTRIBUZIONE BINOMIALE):

Distribuzione binomiale

$$b(k,n;p) = \binom{n}{k} p^k q^{n-k} \text{ dove } q=1-p$$

Il programma BINOMIALE richiede i valori ''p'' e ''n'' e fornisce b(k,n:p) per $k=0,1,\ldots,n$.

Per comodita' di lettura, i valori della tabella vengono presentati con 4 cifre decimali (si tratta della forma tipica utilizzata nelle tavole).

Successivamente si possono richiedere, per uno o piu' k, i valori:
$P[X=k]$ con un miglior grado di approssimazione

$P[X \geqslant k]$ pari a $\sum_{i=k}^{n} P[X=i]$

$P[X \leqslant k]$ pari a $1-P[X \geqslant k]$

Sintesi del listato del programma BINOMIALE:

```
60 INPUT P,N:DIM CB(N),T(N):Q=1-P:M=N:X=1:
   J=0:R=N:S=1:K=0
70 IF M THEN X=X*Q:M=M-1:GO TO 70
80 CB(K)=X:T(K)=U:U=U+X
    :IF K<N THEN X=X*R/S*P/Q:R=R-1:S=S+1:K=
    K+1:GO TO 80
90 FOR K=0 TO N:PRINT K,CB(K),T(K):NEXT K
```

La relazione ricorrente utilizzata per il calcolo di b(k,n;p) e' riportata nel paragrafo 4.6; al termine dell'elaborazione CB(K) contiene $P[X=k]$ e T(K) il valore di $P[X<k]$ per $k=0,1,..,N$.

Con l'aiuto del programma BINOMIALE, risolviamo il problema del compito in classe. Scegliendo $P=0.25$ e $N=10$ otterremo:

K	B(K,N;P)
0	0.0563
1	0.1877
2	0.2816
3	0.2503
4	0.1460
5	0.0584
6	0.0162
7	0.0031
8	0.0004
9	0.0000
10	0.0000

proseguendo con

K = ? 6

si ha:

P[X = 6]=0.0162220001

P[X> = 6]=0.0197277069

P[X < 6]=0.980272293

Cattive notizie per Augusto: la probabilita' di superare la prova (P[X> =6]) e' meno del 2% ... non sempre la fortuna aiuta gli audaci.

Naturalmente, scegliendo opportunamente i valori di N e P, si possono inventare e risolvere diversi problemi legati allo schema delle prove ripetute (esempio: Cosa succede se, ad ogni quesito, vengono proposte 5 alternative - 4 errate + 1 esatta - ? ... si utilizzera' il programma BINOMIALE con $N=10$ e $P=0.2$).

SIMULAZIONE

Nel programma TEST vengono simulate le prove di un gruppo di studenti che si comportano come Augusto.

Inizialmente viene richiesto il numero di studenti (ovvero il numero NS di simulazioni), quindi viene costruito il vettore R%(I) (RISPOSTE) I = 0,1,...,9 che contiene la soluzione del test. Successivamente, per ogni studente, viene calcolato il numero E di risposte esatte (il tentativo dello studente INT(RND(1)*4) + 1 viene confrontato con R%(I) per I = 0,1,...,9).

Il vettore A(E) contiene, in ogni momento, il numero degli studenti che hanno fornito esattamente E risposte corrette (E = 0,1,...,10); il rapporto A(E)/NS esprime la relativa frequenza. Il programma prosegue calcolando la ''frequenza delle prove sufficienti'' ovvero (A(6) + A(7) + A(8) + A(9) + A(10))/NS.

Listato del programma TEST

```
20 DIM R%(9),A(10):INPUT NS
40 FOR I=0 TO 9:R%(I)=INT(RND(1)*4)+1:NEXT I
50 FOR K=1 TO NS:E=0
60 FOR I=0 TO 9:IF INT(RND(1)*4)+1=R%(I) THEN E=E+1
70 NEXT I
80 A(E)=A(E)+1:NEXT K
90 FOR K=6 TO 10::W=W+A(K):NEXT K
95 FOR I=0 TO 10:PRINT I,A(I)/NS:NEXT I:PRINT W/NS
```

Si confrontino i valori ottenuti dalla simulazione con quelli forniti dal programma BINOMIALE per P = 0.25 e N = 10.

Esempio di output del programma TEST

```
SIMULAZIONI ? 70
    K    FREQ.
    0    0.0571428571
    1    0.2
    2    0.242857143
    3    0.257142857
    4    0.171428571
    5    0.0428571429
    6    0.0142857143
    7    0.0142857143
    8    0
    9    0
   10    0

FREQ.SUFF. 0.0285714286
```

2.6 L'ESPLORATORE

Due programmi associati:

"ESPLORATORE 1"
"ESPLORATORE 2"

Un esploratore risulta disperso. Questi si e' recato a ovest con probabilita' Pw, oppure a est con probabilita' Pe = 1-Pw.

Ogni squadra di soccorso, operando nella "zona ovest", lo trova con probabilita' Rw (ammesso che si sia recato da quella parte); operando nella "zona est" lo troverebbe con probabilita' Re.

Le squadre disponibili, che operano indipendentemente l'una dall'altra, sono in numero di "n". Quante inviarne a ovest e quante a est per massimizzare la probabilita' di ritrovamento?

SOLUZIONE

Prima di procedere con i calcoli, e' buona cosa affrontare un esempio in modo da chiarire i termini del quesito e da evidenziarne la "non banalita'" (il problema risulta banale se, ad esempio, Pw = 0; in tal caso tutte le squadre disponibili verranno indirizzate a est; altri casi ovvi si hanno se Re o Rw sono nulli: viene allora esclusa la possibilita' di ritrovamento, in una certa zona, indipendentemente dal numero di squadre operanti).

Supponiamo che sia Pw = 0.6 e Pe = 0.4 (questi potrebbero essere valori da noi stimati in base alle conoscenze delle intenzioni, degli interessi, delle predilezioni dell'esploratore), per cui e' "piu' facile" (ma non di molto) che il soggetto si sia recato a ovest.

Purtroppo a ovest la vegetazione e' assai intricata: una squadra di soccorso ritroverebbe l'esploratore con probabilita' (Rw) di solo 0.1 (se c'e'!).

A est le cose vanno meglio: la visibilita' e' migliore e una squadra lo troverebbe con probabilita' (Re) pari a 0.2.

L'operare indipendente delle squadre significa che la probabilita' di ritrovamento da parte di due differenti squadre e' pari al prodotto delle probabilita' di ritrovamento di ciascuna di esse.

A questo punto dobbiamo decidere come distribuire le squadre, diciamo 13, fra le due zone (vedremo che, nella situazione qui descritta, converra' mandarne 8 a ovest e 5 a est con probabilita' di successo del 61%).

Torniamo al caso generale: sia x il numero delle squadre inviate a ovest ($0 \leqslant x \leqslant n$) e n-x il numero delle squadre inviate a est.

Indichiamo con P(Ax) la probabilita' che l'impresa abbia successo inviando x squadre a ovest. Si calcolera' P(Ax) variando x da 0 a n e si scegliera' il valore di x che rende massima tale quantita'.

Nel nostro caso risulta piu' agevole il calcolo della probabilita' dell'evento Bx, complementare di Ax:

P(Bx) = P[l'esploratore NON viene ritrovato inviando x squadre a ovest (e, quindi, n-x a est)].

Sia H = {l'esploratore si e' recato a ovest},
K = {l'esploratore si e' recato a est};
(H,K) e' una partizione di Ω per cui (vedi 2.1):

$$P(Bx) = P(Bx/H) \cdot P(H) + P(Bx/K) \cdot P(K).$$

Per il calcolo degli addendi, poniamo Fw = 1-Rw; Fw rappresenta la probabilita' di "fallimento nella ricerca" di una singola squadra operante a ovest. Analogamente sia Fe = 1-Re.

Si ha $P(Bx/H) = (Fw)^x$: supponendo che l'esploratore si sia recato a ovest, l'evento Bx si verifica se fallisce sia la squadra 1, sia la squadra 2, sia la 3,..., sia la numero "x" (la funzione vale anche per x = 0: $P(B_0/H) = (Fw)^0 = 1$: se nessuna squadra e' inviata a ovest e l'esploratore si e' recato proprio a ovest, e' certo che non sara' ritrovato). Per l'indipendenza, la probabilita' e' $(Fw)^x$.

Analogamente P(Bx/K) e' $(Fe)^{n-x}$ in quanto tutte le n-x squadre devono fallire.

Poiche' P(H) = Pw e P(K) = Pe si ottiene:

$$P(Bx) = Pw \cdot (Fw)^x + Pe \cdot (Fe)^{n-x}.$$

Soluzione 1

Il numero delle squadre da inviare a ovest puo' essere calcolato senza "scomodare l'analisi"; e' infatti quell'intero x, con $0 \leq x \leq n$, che rende minima P(Bx).

Il programma ESPLORATORE 1 richiede:

- il numero di squadre a disposizione (N:N > 1)
- la probabilita' che l'esploratore si sia recato a ovest (PW: $0 \leq PW \leq 1$)
- la probabilita' che una singola squadra operante nella zona ovest riesca a localizzarlo (RW: $0 \leq RW \leq 1$)
- come sopra, per la zona est (RE: $0 \leq RE \leq 1$)

determina:
la ripartizione delle n squadre che massimizza la probabilita' di ritrovamento e il valore di tale probabilita' (su richiesta, vengono stampate le probabilita' di ritrovamento per le altre ripartizioni nel formato "numero di squadre a ovest / probabilita' di successo").

Listato del programma ESPLORATORE 1

```
10 INPUT''N.SQUADRE'';N
```

```
 30 INPUT''PROB.ESPL. A OVEST'';PW
 55 INPUT''PROB.RITROVAMENTO SQUADRA IN
    OVEST'';RW
 75 INPUT''PROB.RITROVAMENTO SQUADRA IN EST '';RE
100 PE=1-PW:FW=1-RW:FE=1-RE:MN=2
110 FOR X=0 TO N
120 PB=PW*FW↑X+PE*FE↑(N-X):PRINT X,1-PB
140 IF PB<MN THEN Y=X:MN=PB
150 NEXT X
160 PRINT''SQUADRE A OVEST'';Y
170 PRINT''SQUADRE A EST'';N-Y
180 PRINT''PROB. DI SUCCESSO'';1-MN
```

Nel ciclo principale (istr. 110-150), P(Bx) viene calcolato per tutte le possibili ripartizioni (X rappresenta il numero di squadre inviate a ovest); quando, per un certo X, la probabilita' di fallimento e' minore di quella ottenuta nei casi precedenti, il valore di X viene caricato in Y. A ciclo concluso, Y contiene la soluzione.

Soluzione 2

Il problema puo' anche essere risolto con "un po' di analisi". Sia $k = Pw$; $h = Pe$; $a = Fw$; $b = Fe$; si tratta allora di determinare il minimo di

$$f(x) = k \cdot a^x + h \cdot b^{n-x} \quad \text{per } x \epsilon [0;n]$$
$$\text{con } k + h = 1 \ .$$

Se consideriamo a, b, h, k strettamente compresi tra 0 e 1, si ha:
$f'(x) = k \cdot a^x \cdot \log a - h \cdot b^{n-x} \cdot \log b$
per cui $f'(x) \geqslant 0$ se $k \cdot a^x \cdot \log a \geqslant h \cdot b^{n-x} \cdot \log b$.

Poiche' $a < 1$ e $k > 0$, e' $k \cdot \log a < 0$ mentre $b^{n-x} > 0$; dividendo per $b^{n-x} \cdot k \cdot \log a$ (<0), si ha:

$$\frac{a^x}{b^{n-x}} \leqslant \frac{h \cdot \log b}{k \cdot \log a}$$

Posto $M = (h \cdot \log b)/(k \cdot \log a)$, passando al logaritmo di entrambi i membri, e':

$x \cdot \log a - (n-x) \cdot \log b \leqslant \log M$
ovvero
$x(\log a + \log b) \leqslant \log M + n \cdot \log b$.

Poiche' $\log a + \log b < 0$ (essendo $a,b \epsilon (0,1)$), e'

$$x \geqslant \frac{\log M + n \cdot \log b}{\log a + \log b} = x'$$

In definitiva, per $x = x'$ si ha un minimo di $f(x)$.

Dato che $f(x)$ e' definita e continua per ogni x reale e x' e' l'unico minimo relativo, volendo limitare il calcolo del minimo in [0,n], si avra'

se $0 \leqslant x' \leqslant n$: minimo in x'
se $x' < 0$: minimo in 0
se $x' > n$: minimo in n.

Visto che interessa un valore intero, prenderemo l'intero piu' vicino a x' (istr. 150).

Nel programma ESPLORATORE 2, la soluzione viene trovata in questo modo senza che sia necessario passare in rassegna le varie ripartizioni come avveniva nel programma precedente.

Listato del programma ESPLORATORE 2
(le variabili N, PW ecc. corrispondono alle analoghe di ESPLORATORE 1)

```
100 INPUT N,PW,RW,RE:PE=1-PW:FW=1-RW:FE=1-RE
110 M=PE*LOG(FE)/(PW*LOG(FW))
120 Y1=(LOG(M)+N*LOG(FE))/(LOG(FW)+LOG(FE))
130 IF Y1<0 THEN Y1=0
140 IF Y1>N THEN Y1=N
150 Y1=INT(Y1+.5):PF=PW*FW↑Y1+PE*FE↑(N-Y1)
160 PRINT Y1,N-Y1,1-PF
```

Esempi

esempio 1

Numero squadre	10
Prob. esploratore a W	0.2
Prob. ritrovamento a W	0.9
Prob. ritrovamento a E	0.2

Conviene inviare: 1 squadra a ovest
9 squadre a est

la probabilita' di successo e' circa 0.87

esempio 2

Numero squadre	71
Prob. esploratore a W	0.45
Prob. ritrovamento a W	0.04
Prob. ritrovamento a E	0.047

Conviene inviare: 34 squadre a ovest
37 squadre a est

la probabilita' di successo e' circa 0.795

2.7 IL DADO TRUCCATO

Un programma associato:

''BAYES''

Un sacchetto contiene due dadi:
- uno ''regolare'': uscite da 1 a 6 equiprobabili
- uno ''truccato'': lanciandolo si ottiene sempre 6.

Un giocatore ha estratto uno dei dadi e, senza esaminarlo, l'ha tirato (una volta) ottenendo 6.

Qual e' la probabilita' che abbia utilizzato il dado truccato?

SOLUZIONE

Posto H = {e' stato lanciato il dado truccato}
K = {e' stato lanciato il dado regolare}
A = {e' uscito il 6},
e' $H \cup K = \Omega$ e $H \cap K = \emptyset$.

Vogliamo calcolare P(H/A), cioe' la probabilita' che sia stato utilizzato il dado truccato, SAPENDO che e' uscito il 6.

Si tratta di una semplice applicazione del teorema di Bayes. Nel nostro esempio e':

$P(A) = P(A/H) \cdot P(H) + P(A/K) \cdot P(K)$
e $P(H/A) \cdot P(A) = P(A/H) \cdot P(H) = P(A \cap H)$
quindi

Formula di Bayes

$$P(H/A) = \frac{P\,(A/H) \cdot P\,(H)}{P\,(A/H) \cdot P(H) + P\,(A/K) \cdot P(K)}$$

Poiche' $P(A/H) = 1$; $P(H) = P(K) = 1/2$; $P(A/K) = 1/6$, si ottiene $P(H/A) = 6/7 = 0.857142..$.

Un confronto tra P(H) e P(H/A) mostra quanto l'ipotesi H venga avvalorata dall'informazione ''e' uscito il 6'' (si passa dal 50% a circa l'86%).

SIMULAZIONE

Nel breve programma che segue, le variabili T e B fungono da contatori:

T conta quante volte viene estratto il dado truccato

B conta il ''numero di 6'' ottenuti nelle N prove

(''A'' rappresenta il valore uscito nella singola prova).

Il rapporto T/B fornisce la frequenza di utilizzo del dado truccato nei lanci che hanno dato un 6. Nei termini della legge dei grandi numeri, T/B approssimera' P(H/A).

Listato del programma BAYES

```
10 H=.5:U=1:S=6:INPUT''N. PROVE'';N
20 FOR I=U TO N:IF RND(U)<H THEN T=T+U:A=
   S:GO TO 40
30 A=INT(RND(U)*S)+U
40 IF A=S THEN B=B+U
50 NEXT I
60 PRINT''FREQ ='';T/B
```

Esempio di output del programma Bayes

```
* BAYES *
N.PROVE 1000
FREQ. =0.847058824
[Il valore teorico previsto e' 6/7=0.857142]
```

2.8 CALCOLIAMO IL LOGARITMO

Un programma associato:

''LOGARITMO''

Metodo Monte Carlo

Le tecniche matematiche che consentono la determinazione di un certo valore ''x'' attraverso un un metodo aleatorio portano il nome di metodi Monte Carlo. In [3] viene proposto l'esempio classico del calcolo di π; con procedimento analogo viene qui mostrato come valutare il logaritmo naturale di un numero maggiore di uno.

In un sistema di assi cartesiani, consideriamo la funzione $y=1/x$ nell'intervallo $[1,b]$ con $b>1$.

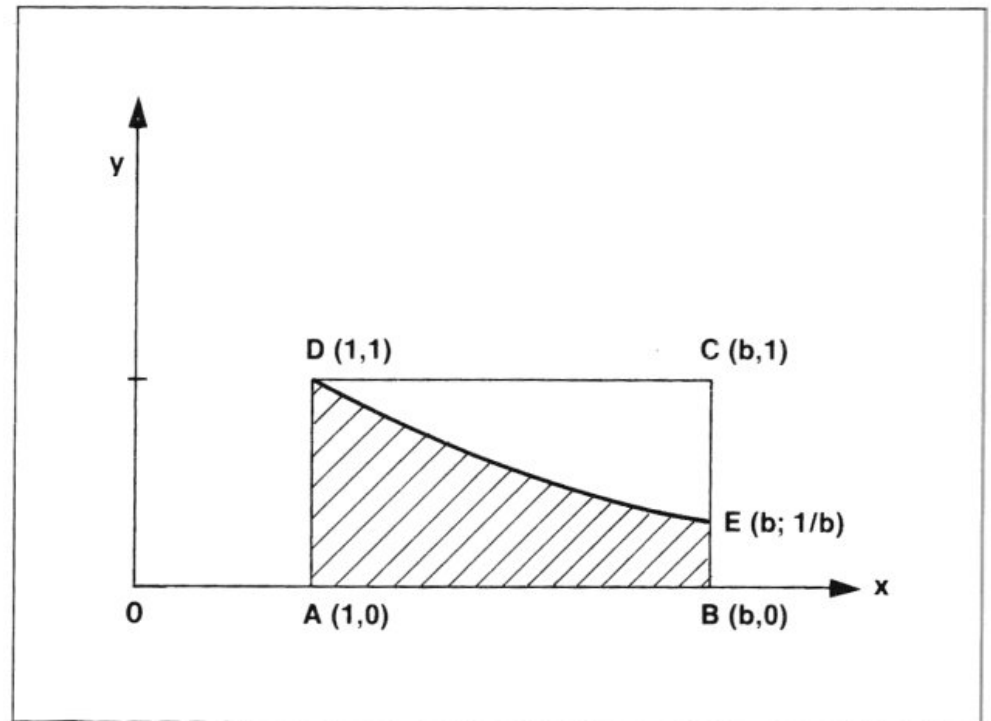
Sia ABCD il rettangolo di vertici $A=(1,0)$; $B=(b,0)$; $C=(b,1)$; $D=(1,1)$ e ABED il trapezoide limitato dall'asse x, dal grafico di $1/x$ e dalle rette $x=1$ e $x=b$ (quindi $E=(b, 1/b)$ - vedi fig. 6 -).

La probabilita' che un punto P, scelto a caso nel rettangolo, appartenga anche al trapezoide e' data dal rapporto delle aree delle due figure:

p = S(trapezoide ABED)/S(rettangolo ABCD).

Si ha:

$$\text{S(trapezoide ABED)} = \int_1^b \frac{dx}{x} = \log b$$

Fig. 6 *Calcolo di log b con il metodo Montecarlo*

S(rettangolo ABCD) = (b - 1)·1 = b - 1

quindi p = (log b)/(b-1) ovvero log b = (b-1)p

SIMULAZIONE

Nel programma che segue, all'istruzione 10 viene richiesto il numero B di cui si vuole calcolare il logaritmo e il numero di simulazioni (NS).

Il ciclo principale e' costituito dalle istruzioni 30-50: ad ogni iterazione viene scelto casualmente un punto di coordinate X,Y appartenente al rettangolo (RND(U) fornisce un valore pseudocasuale tra 0 e 1); alla linea 40 si controlla se tale punto appartiene anche al trapezoide, nel qual caso si incrementa il contatore S.

Per NS sufficientemente elevato, S/NS e' una buona approssimazione di p (nei termini della legge dei grandi numeri) e V = (B-1)*S/NS e' il valore approssimato del logaritmo naturale di B (istr. 60).

Il programma si conclude con la stampa del valore simulato (V), del valore fornito dalla macchina (LB), dell'errore assoluto e di quello relativo rispetto a LB.

Listato del programma LOGARITMO

```
10 INPUT B,NS:U=1:C=B-U:LB=LOG(B)
30 FOR I=U TO NS:Y=RND(U):X=C*RND(U)+U
40 IF Y<=U/X THEN S=S+U
50 NEXT I
60 V=C*S/NS:EA=ABS(V-LB):PRINT V,LB,EA,EA/LB
```

Esempio di output del programma LOGARITMO

NUM. (> 1)	: 24
NUM. SIM.	: 1000
SIMULATO	: 3.243
LOG(B)	: 3.17805385
ERR. ASS.	: 0.0649
ERR. REL.	: 0.0204

(Il rapido appiattimento di $1/x$ per elevati valori della variabile e' responsabile delle scarsa attendibilita' dei risultati per ''grandi'' valori di b; qui si innesterebbe un discorso sulla precisione-macchina che esula da questa sede).

2.9 L'EQUAZIONE ALEATORIA

Un programma associato:

''EQUAZIONE''

I coefficienti a, b, c dell'equazione $ax^2+bx+c=0$ sono numeri reali scelti ''a caso'' in un intervallo.

Qual e' la probabilita' che l'equazione abbia soluzioni reali?

SOLUZIONE

Supponiamo $a \neq 0$ [e' $P(a=0)=0$]. Senza perdere di generalita', possiamo trasformare l'equazione in $x^2+dx+c=0$ dove ''d'' e ''c'' sono due reali scelti a caso.

L'equazione fornisce soluzioni reali se $d^2-4c \geqslant 0$ ovvero se:

$$c \leqslant d^2/4 \qquad (*).$$

La probabilita' che la coppia (d,c) di $\mathbb{R}^2$ soddisfi la (*) e' pari alla probabilita' che un punto $P=(x,y)$, scelto a caso nel piano cartesiano, sia situato ''al di sotto'' del grafico di $y=x^2/4$ (oppure sul grafico).

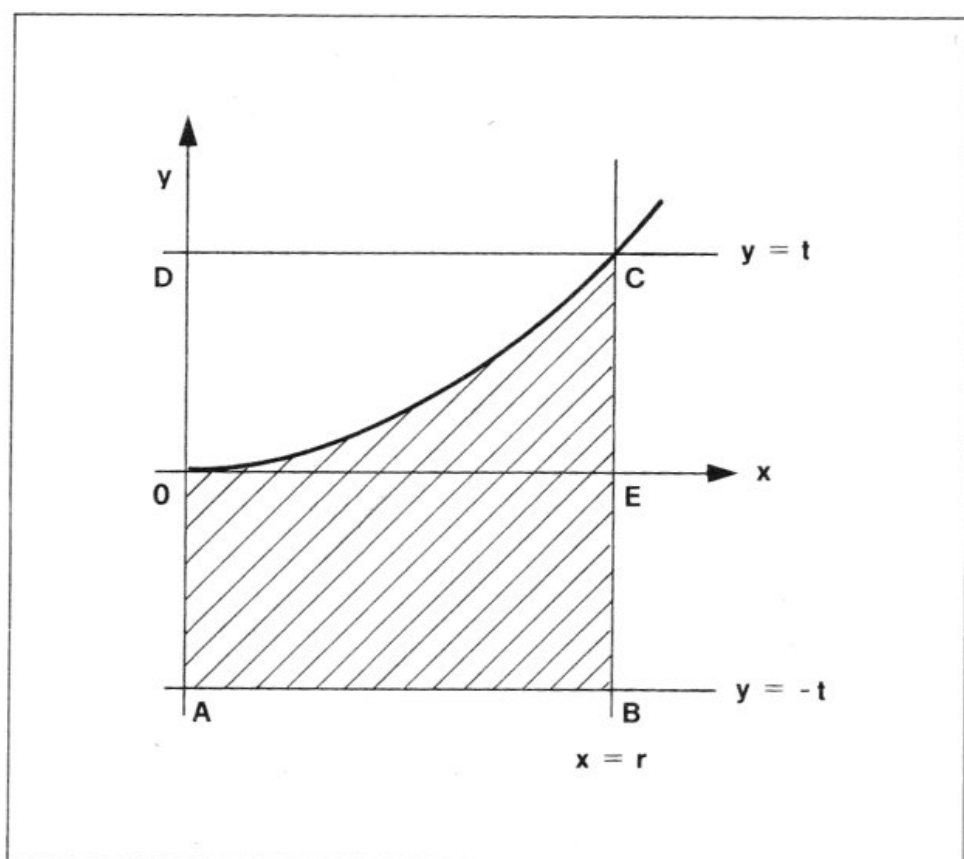
Per ragioni di simmetria, bastera' considerare il caso $x \geqslant 0$.

Sia r un numero reale positivo e $t=r^2/4$. Scegliamo a caso un punto P appartenente al rettangolo R limitato dall'asse y, dalla retta $x=r$ e dalle rette $y=t$ e $y=-t$ (vedi fig.7).

La probabilita' P(r) che P sia situato ''al di sotto'' dell'arco OC e' pari a

Area(trapezoide OABC)/Area(rettangolo ABCD) = A(T)/A(R).

E'

Fig. 7 *Probabilità "geometrica" (problema 2.9)*

$$A(T) = \frac{r^3}{4} + \int_0^r \frac{x^2}{4}\,dx = \frac{1}{3}r^3$$

$$\text{e } A(R) = \frac{1}{2}r^3$$

quindi $P(r) = 2/3$.

Essendo ovviamente $\lim_{r \to \infty} P(r) = 2/3$, 2/3 e' la probabilita' richiesta.

SIMULAZIONE

Nel programma EQUAZIONE vengono richiesti due valori: K e N. Per un fissato valore di K (>1), il programma determina "casualmente" i tre coefficienti dell'equazione nell'intervallo (-K, +K) e, per ciascuna delle N simulazioni, controlla se l'equazione aleatoria fornisce o meno soluzioni reali.

Per N sufficientemente elevato, ci si attende una frequenza prossima a 2/3; la frequenza ottenuta viene fornita al termine dell'elaborazione in forma di rapporto (numero di "successi"/numero prove simulate) e in forma decimale.

Dal punto di vista teorico, il valore K non e' essenziale (si e' visto che $P = 2/3$ per ogni $K > 0$), ma puo' rivelarsi importante in relazione alla precisione-macchina.

Listato del programma EQUAZIONE

```
30 INPUT''INTERVALLO (>1)'';K
35 DEF FNF(X)=2*K*RND(1)-K:INPUT''NUM.SIMUL.'';N
60 FOR I=1 TO N:A=FNF(1):B=FNF(1):C=FNF(1):D=B*B-4*A*C:IF
   D>= 0 THEN R=R+1
70 NEXT I:PRINT''FREQ.='';R;''/'';N:PRINT''FREQ.='';R/N
```

Esempio di output del programma EQUAZIONE

INTERVALLO (>1)? 1000
NUM. SIMULAZIONI? 300
FREQ. = 185/300
FREQ. = 0.616666667
[P=2/3=0.667 CIRCA]

2.10 PING PONG 1

Un programma associato:

''PING PONG 1''

In un certo gioco, Aldo ''guadagna un punto'' contro Bruno con probabilita' fissata ''p''.

Vince la gara chi per primo raggiunge un vantaggio di due punti sull'altro.

Calcolare:

1. la probabilita' che Aldo si aggiudichi la gara supponendo che le succcessive realizzazioni dei vari punti siano eventi mutualmente indipendenti;
2. il numero medio di ''giochi'' necessari alla conclusione della partita (somma finale dei punteggi ottenuti da Aldo e Bruno).

Questa ''regola di torneo'' viene utilizzata nella parte finale di giochi quali il tennis e il tennis da tavolo.

Esempio:

Azione	Situazione	Punti A.	Punti B.
--	parita'	0	0
punto per A	vantaggio A	1	0
punto per B	parita'	1	1
punto per B	vantaggio B	1	2
punto per B	FINE, vince B	1	3

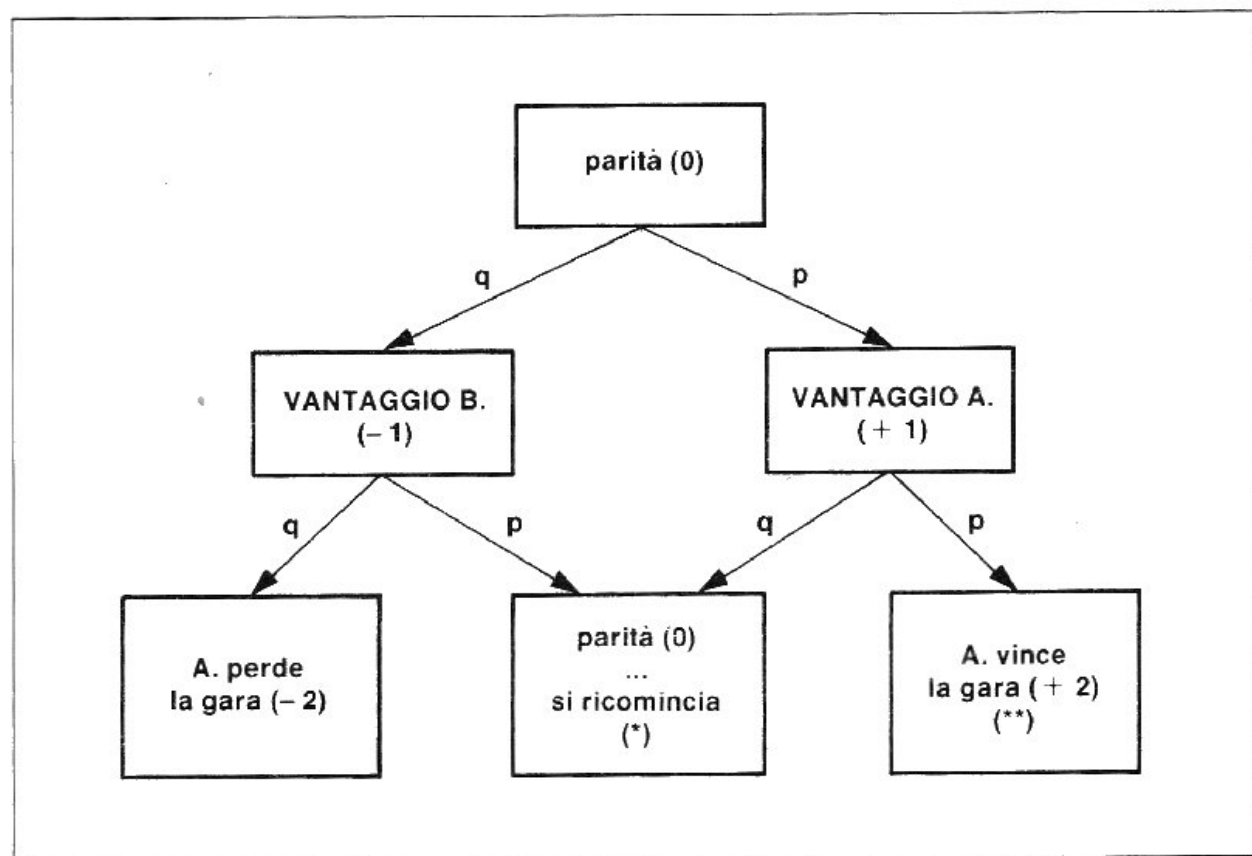



Fig. 8 *Grafo per la soluzione del problema 2.10*

La gara si e' conclusa con 4 ''giochi''.

SOLUZIONE

Sia x la probabilita' richiesta, ovvero
x = P[Aldo vince la gara] =
= P[Aldo vince con situazione iniziale di parita']
e sia q = 1-p = P[Aldo perde un punto].

L'andamento del gioco puo' essere rappresentato dallo schema di fig.8.

E'

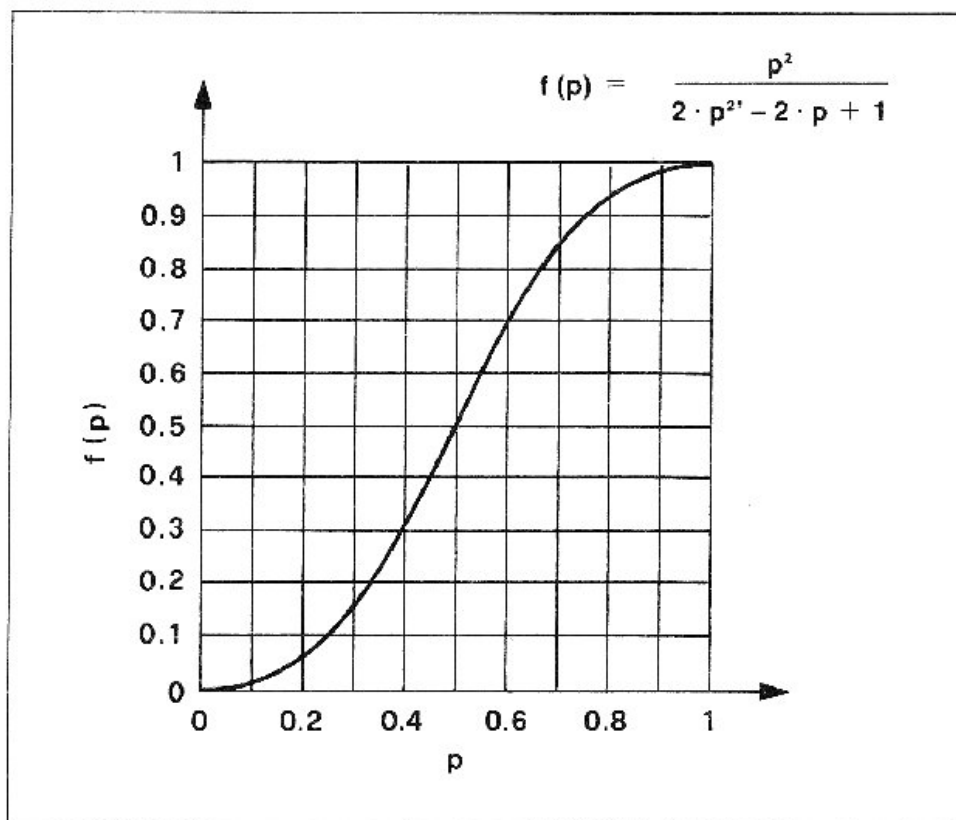
$$x = p^2 + 2pqx$$

(p^2 e' la probabilita' di raggiungere la situazione (**); 2pq e' la probabilita' di trovarsi nella situazione (*): da li' la probabilita' di aggiudicarsi la gara e' nuovamente x).

Ricavando x si ottiene:

$$x = f(p) = \frac{p^2}{1 - 2pq} = \frac{p^2}{2p^2 - 2p + 1}$$

Lo studio di f(p) in [0,1] (vedi fig.9), mostra come il meccanismo di gioco amplifichi la probabilita' di vittoria della gara quan-

Fig. 9 *Probabilità di vincere la gara in funzione della probabilità di guadagnare il punto (problema 2.10)*

do la probabilita' di guadagnare il punto sia maggiore di 0.5; questo significa che la ''regola di torneo'' contribuisce ad evidenziare il migliore:

p	f(p)
0.5	.50
0.6	0.69
0.7	0.84
0.8	0.94
0.9	0.99

Durata media della gara

Sia D la ''durata media della partita'' ovvero il numero medio di prove neccessarie a chiudere la gara. Evidentemente e' $D \geqslant 2$, ricorrendo allo schema di fig.8, si ha:

$$D = 2(p^2 + q^2) + (D + 2)2pq$$

(E' $D = 2$ se Aldo o Bruno realizzano subito 2 punti consecutivi, cio' avviene con probabilita' $p^2 + q^2$; con probabilita' 2pq si giunge ad una situazione di parita': servono allora, mediamente, altre D prove per concludere).

Ricavando D si ottiene:

$$D = g(p) = \frac{2}{1 - 2pq} = \frac{2}{2p^2 - 2p + 1}$$

con $0 \leqslant p \leqslant 1$.

Come prevedibile, $g(0) = g(1) = 2$ (se uno dei giocatori e' imbattibile, la partita dura esattamente 2 giochi) e $g(1/2 + t) = g(1/2 - t)$ per $t \in [0,1/2]$ (uno ''scambio di abilita''' tra Aldo e Bruno non altera la durata della gara).

Derivando, si scopre che $g'(p) \geqslant 0$ per $p \leqslant 1/2$; la funzione ammette un massimo nel punto $M = (1/2,4)$: una partita tra giocatori di pari forza dura di piu', mediamente 4 giochi.

Il programma PING PONG 1 richiede P e fornisce f(P) e g(P); le istruzioni dalla 50 alla 90 realizzano la simulazione.

Per ogni simulazione ($I = 1,2,...,N$), viene generato un valore pseudocasuale R dell'intervallo (0,1).

Se $R < P$, Aldo guadagna un punto; in caso contrario lo guadagna Bruno.

La variabile S rappresenta il vantaggio relativo di Aldo su Bruno; quando $S*S = 4$ la partita ha termine ed il totale delle vittorie di Aldo (T) viene incrementato se necessario; in questo caso si ripristina la situazione di parita' e si procede ad una nuova simulazione (istr. 90).

Se $S*S < 4$, la gara non e' terminata: si rientra alla 60 per l'assegnazione di un nuovo punto; la variabile W conta il numero totale dei giochi effettuati.

Alla 92 viene stampata la frequenza di vittoria di Aldo (numero di gare vinte T sulle N simulate) da confrontare con il valore teorico f(P); alla 94, operazione analoga per il numero medio di giochi per partita (W/N) da confrontare con g(P).

Listato del programma PING PONG 1

```
30 U=1:F=4:D=2:Z=0:INPUT''PROB. DI GUADAGNARE IL PUNTO
   '';P
40 Q=1-P:PRINT''PROB. DI VINCERE LA GARA'';P*P/(1-2*P*Q)
47  PRINT''NUM.  MEDIO  TIRI  PER  GARA  (PRE-
VISTO)'';2/(1-2*P*Q)
50 INPUT''NUM. SIMUL.'';N
55 FOR I=U TO N
60 W=W+U:R=RND(U):S=S-(R<P)+(R>=P)
70 IF S*S=F THEN 90
80 GO TO 60
90 T=T-(S=D):S=Z:NEXT I
92 PRINT''FREQUENZA'';T/N
94 PRINT''NUM. MEDIO TIRI PER GARA (OTTENUTO)'';W/N
```

Esempio di output

Esempio di output del programma PING PONG 1

PROB. DI GUADAGNARE IL PUNTO ? 0.75

PROB. DI VINCERE LA GARA	0.9
NUM. MEDIO TIRI PER GARA (PREVISTO)	3.2
NUM. SIMULAZIONI ?	517
FREQUENZA	0.899419729
NUM. MEDIO TIRI PER GARA (OTTENUTO)	3.30754352

2.11 PING PONG 2

Un programma associato:
"PING PONG 2"

E' noto che nel gioco del tennis (anche nel tennis da tavolo, nella pallavolo ecc.), e' preferibile trovarsi "alla battuta" (o "in possesso di palla") piuttosto del contrario ed e' realistico pensare che la probabilita' dell'evento

{il giocatore X guadagna un punto}

sia maggiore quando il servizio tocca a X.

Riprendiamo il problema proposto in PING PONG 1: il regolamento e' lo stesso, ma Aldo e Bruno si alternano alla battuta.

Aldo inizia: p e' la sua probabilita' di guadagnare il punto quando e' in possesso di palla; p' l'analoga probabilita' nell'altro caso.

Anche qui si vuol calcolare:

1. la probabilita' che Aldo vinca la gara
2. la durata media della gara.

OSSERVAZIONE Se $p = p'$ si ritorna alla situazione descritta in PING PONG 1; converra' supporre $p > p'$. Per $p < p'$ si prospetta una situazione anomala, in cui Aldo gioca meglio quando non e' in possesso di palla.

SOLUZIONE

Detto "successo" la realizzazione di un punto da parte di Aldo, la situazione descritta determina uno schema delle prove ripetute di parametro p nelle prove dispari e parametro p' nelle prove pari.

Posto $q = 1-p$ e $q' = 1-p'$, mantenendo le notazioni adottate in PING PONG 1, si ha lo schema di fig.10.

Se

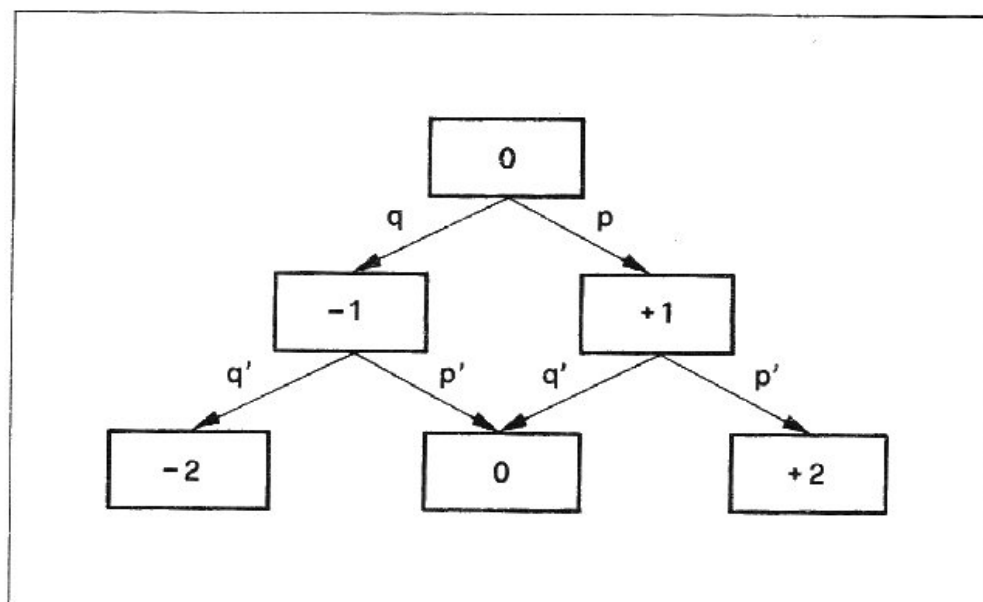
x = P[Aldo vince la gara]

e'

$$x = pp' + (pq' + qp')x$$

da cui:

$$x = \frac{pp'}{1 - pq' - qp'} = \frac{pp'}{2pp' - (p + p') + 1}$$

Fig. 10 *Schema relativo al problema 2.11*

Se D e' la durata media della gara, e'

$D=2(pp' + qq') + (D + 2)(pq' + qp')$

da cui:

$$D = \frac{2}{1 - pq' - qp'} = \frac{2}{2pp' - (p + p') + 1}$$

(Se p = 1 e p' = 0, o viceversa, la partita non ha fine: il denominatore si annulla e x e D perdono di significato).

Il programma PING PONG 2 richiede i valori p e p' (variabili P e PP), fornisce x e D e realizza la simulazione in modo analogo a quanto avveniva in PING PONG 1. Nei vari giochi, la variabile L assume alternativamente i valori +1 e -1: quando L = 1 la probabilita' di successo e' P, quando L = -1 la probabilita' di successo e' PP.

Listato del programma PING PONG 2

```
10 U=1:F=4:D=2:Z=0:INPUT P,PP
40 Q=1-P:QQ=1-PP:PRINT''x='';P*PP/(1-P*QQ-Q*PP)
45 PRINT''D='';2/(1-P*QQ-Q*PP)
50 INPUT''NUM. SIMUL.'';N:FOR I=U TO N:L=-U
60 L=-L:W=W+U:R=RND(U)
62 IF L=U THEN S=S-(R<P)+(R>=P)
64 IF L=-U THEN S=S-(R<PP)+(R>=PP)
70 IF S*S=F THEN 90
80 GO TO 60
90 T=T-(S=D):S=Z:NEXT I
95 PRINT''FREQUENZA'';T/N
97 PRINT''NUM. MEDIO TIRI PER GARA (OTTENUTO)'';W/N
```

Fig. 11 *Probabilità di vincere la gara al variare delle probabilità di guadagnare il punto della battuta p e alla risposta p' (problema 2.11)*

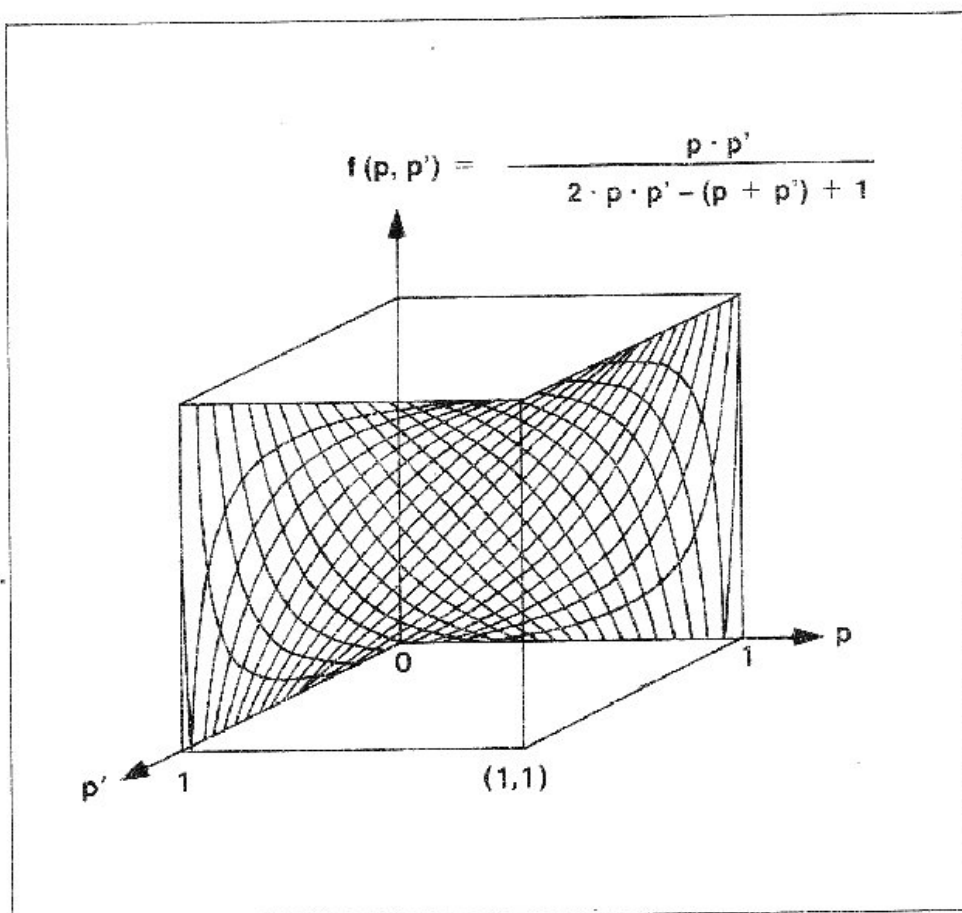
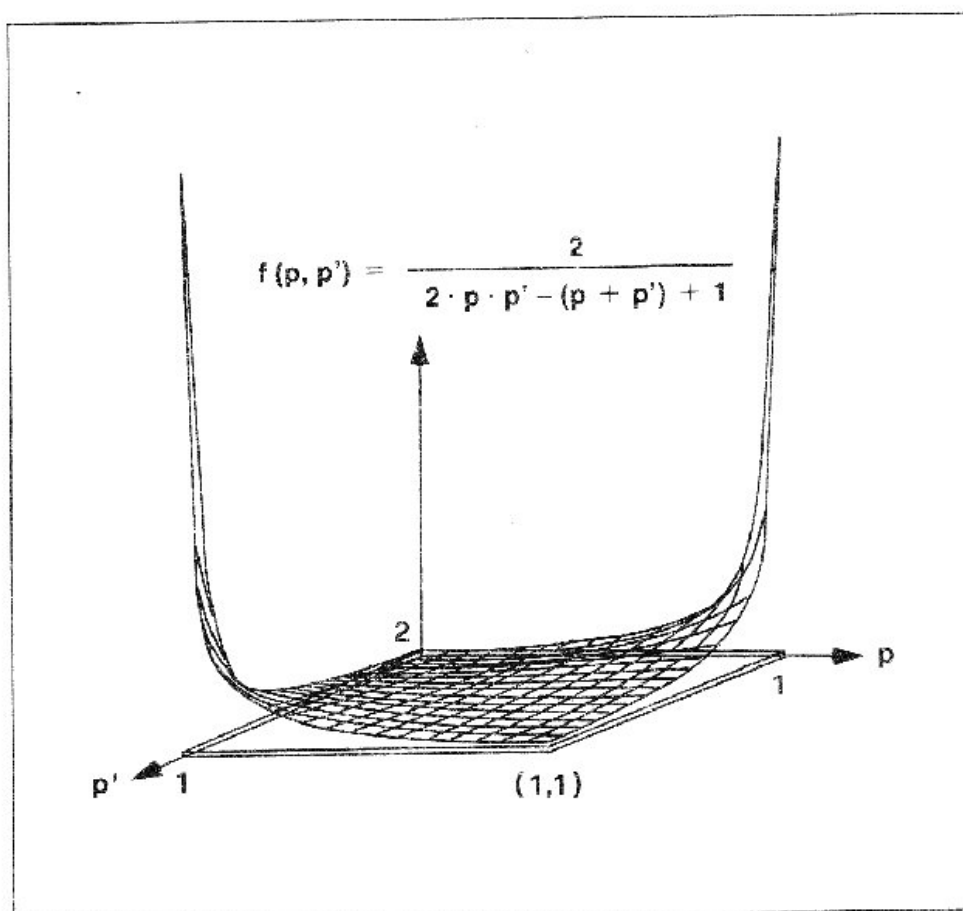



Fig. 12 *Durata media della gara (problema 2.11)*

Esempio di output del programma PING PONG 2

```
  PROB. DI GUADAGNARE IL PUNTO
- ALLA BATTUTA ?  0.7
- ALLA RISPOSTA ?  0.4
  PROB. DI VINCERE LA GARA :                          0.609
  NUM. MEDIO DI TIRI PER GARA (PREVISTO)            :4.348
*SIMULAZIONE*
  NUM. SIMULAZIONI ? 500
  FREQUENZA :                                         0.596
  NUM. MEDIO DI TIRI PER GARA (OTTENUTO) :  4.308
```

Se p e' prossimo a 1 e p' prossimo a 0, Aldo, di norma, passa in vantaggio quando e' al servizio e perde il vantaggio subito dopo. La durata media della gara diviene elevata e la simulazione richiede tempo (per $p=0.99$ e $p'=0.01$ il numero medio di tiri per gara e' 101.01).

2.12 BOWLING

Un programma associato

"BOWLING"

Ester sta concludendo una partita di bowling: due sono i birilli che le restano da abbattere.

Da precedenti statistiche e' noto che Ester, con un tiro, riesce ad abbattere:

- 2 birilli con probabilita' 0.1
- 1 birillo con probabilita' 0.4
- nessun birillo con probabilita' 0.5

quando i birilli rimasti sono 2;

- 1 birillo con probabilita' 0.4
- nessun birillo con probabilita' 0.6

quando resta da colpire l'ultimo birillo.

Quanti tiri, mediamente, occorreranno ad Ester per terminare la partita?

SOLUZIONE

La situazione puo' essere rappresentata attraverso lo schema di

Fig. 13 *Grafo per la soluzione del problema 2.12*

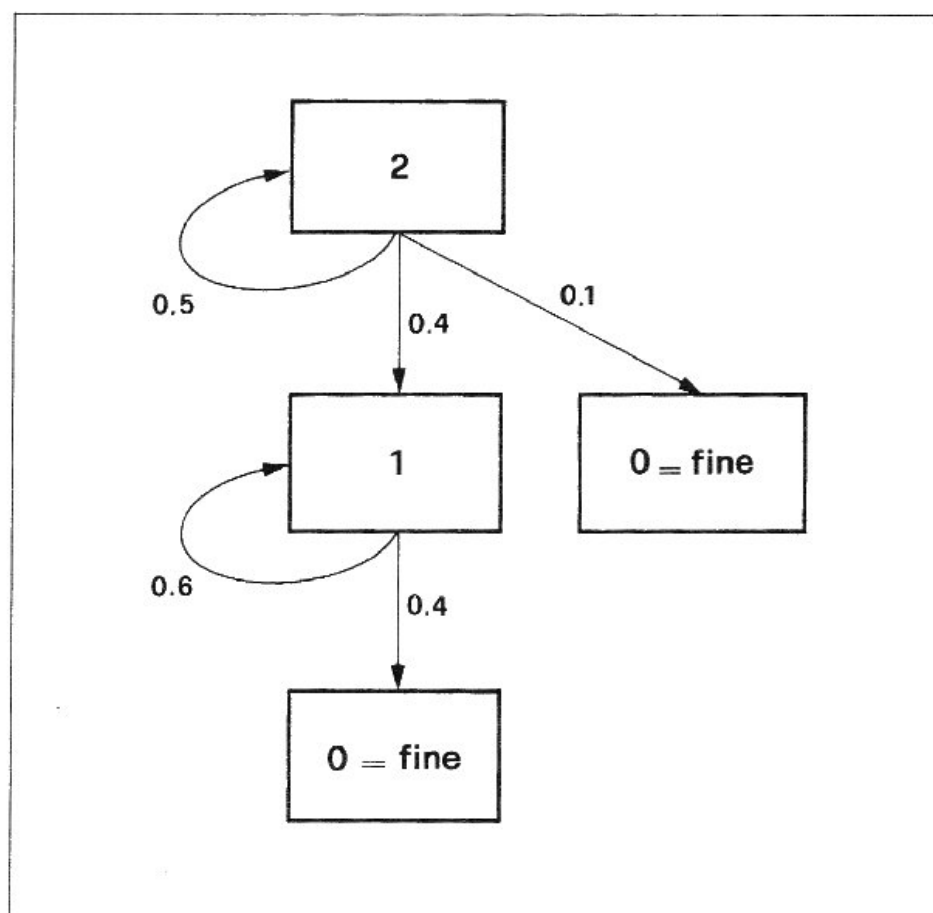



fig. 13 (il numero nella casella rappresenta il numero di birilli che restano).

Sia X il numero medio di tiri che occorrono per completare la partita quando 2 sono i birilli da abbattere; Y l'analogo valore quando resta un solo birillo.

Si ha:

$$\begin{cases} Y = 0.4*1 + 0.6*(1 + Y) \\ X = 0.1*1 + 0.4*(1 + Y) + 0.5*(1 + X) \end{cases}$$

infatti:

- se resta da abbattere un solo birillo, bastera' un tiro con probabilita' 0.4; con probabilita' 0.6 il tiro viene ''sprecato'' e, da li', serviranno altri Y tiri; in totale (1 + Y) (equazione n. 1)
- se i birilli da abbattere sono 2, questi verranno eliminati con un tiro con probabilita' 0.1; con probabilita' 0.4 ne verra' abbattuto 1 e ne serviranno, mediamente, altri Y (quindi 1 + Y); con probabilita' 0.5 il tiro ''va a vuoto'': ne serviranno altri X (equazione n. 2).

Risolvendo il sistema, si ottiene $X = 4$, $Y = 10/4$: ad Ester occorreranno mediamente 4 tiri per terminare la partita.

SIMULAZIONE

Il programma BOWLING richiede il numero N di simulazioni e fornisce il numero medio di tiri che si sono resi necessari per completare ciascuna delle N partite.

Per ogni partita ($I = 1,2,...,N$), il numero di birilli B e' inizialmente posto uguale a 2, mentre viene azzerato il numero di tiri T (istruzione 60).

Nell'istruzione 70 viene originato un numero casuale X dell'intervallo (0,1):
- se $B = 2$ e $X > 0.5$, B rimane 2 (nessun birillo abbattuto)
- se $B = 2$ e $0.1 < X \leqslant 0.5$, B diviene 1 (1 birillo abbattuto)
(segue che, se $B = 2$ e $X \leqslant 0.1$, B diviene 0: tutti abbattuti)
- se $B = 1$ e $X < 0.6$, B rimane 1 (non abbattuto)
(segue che, con $B = 1$ e $X \geqslant 0.6$, B diviene 0: abbattuto).

Finche' $B \neq 0$, la partita non e' terminata, i tiri si susseguono (istruzione 70) e il loro numero viene opportunamente incrementato: $T = T + 1$. Quando $B = 0$, viene aggiornata la variabile S (totale dei tiri in tutte le simulazioni) e il programma procede ad una nuova simulazione ripristinando la situazione iniziale: $B = 2$ e $T = 0$.

Al termine, S/N fornisce il valore cercato.

Listato del programma BOWLING

```
50 INPUT N
60 FOR I=1 TO N:B=2:T=0
70 X=RND(1):B=(B=2)*(2*(X>.5)+(X>.1 AND
X<=.5))+(B=1)*(X<.6):T=T+1:IF B<>0THEN 70
90 S=S+T:NEXT I:PRINT S/N
```

Esempio di output del programma BOWLING

QUANTE PARTITE VUOI SIMULARE ? 379
NUMERO MEDIO OTTENUTO:

3.92084433 TIRI
PER PARTITA.

2.13 DIVERSE USCITE

Due programmi associati:

''I TRE AMICI''
''I SEI AMICI''

Un giocatore lancia un dado regolare una o piu' volte. Se, al termine del gioco, avra' ottenuto uscite tutte diverse, ricevera' un numero di dollari pari alla somma delle uscite stesse; nulla in caso contrario.

Come deve comportarsi:

(a) nel caso possa decidere di fermarsi quando lo desidera

(b) nel caso debba dichiarare, prima di iniziare a giocare, il numero di lanci che intende eseguire.

SOLUZIONE (a)

Si supponga di aver eseguito con successo ''n'' lanci e siano $u_1, u_2, \ldots, u_n$ i valori ottenuti ($u_i \neq u_j$ se $i \neq j$).

Vogliamo considerare l'opportunita' di procedere con un ulteriore lancio (sara' $1 \leqslant n \leqslant 5$; e' ovvio che, se $n = 6$, ci si ferma guadagnando 21$).

Sia $A = \{1,2,3,4,5,6\}$ (valori possibili)
$U = \{u_1, u_2, .., u_n\}$ (valori usciti)
$D = A \setminus U$ (valori disponibili)

Se ci si ferma, si guadagna la somma $s = \sum_{x \in U} x = \sum_{i=1}^{n} u_i$;

se si prosegue, si guadagna: $s + x$ se $x \in D$
0 se $x \in U$.

L'insieme D e' formato da 6-n elementi e

$$\sum_{x \in D} x = \sum_{x \in A} x - \sum_{x \in U} x = 21 - s.$$

Il guadagno atteso, decidendo di continuare, e' dato da:

$$G = \sum_{x \in D} \frac{1}{6} (s + x) = \frac{1}{6} \left[\sum_{x \in D} s + \sum_{x \in D} x \right] =$$

$$= \frac{1}{6} [s(6\text{-}n) + 21\text{-}s] =$$

$$= \frac{1}{6} [5s\text{-}ns + 21] \qquad (*)$$

Converra' continuare a giocare se $G > s$ ovvero se

$$5s - ns + 21 > 6s$$
$$s + ns < 21$$
$$s < \frac{21}{n + 1} \quad (**)$$

Calcolando l'espressione per i diversi valori di n, si ottiene il seguente ''piano di gioco''. Nelle colonne ''min s'' e ''Max s'' sono riportati il minimo e il massimo valore di s per il corrispondente n.

Un confronto tra questi valori e quelli ottenuti dalla (**) consente di raggiungere le conclusioni espresse nell'ultima colonna:

lanci effettuati	min s	Max s	conviene proseguire se ''s'' e' minore di ...
1	1	6	10.5 ... sempre
2	3	11	7.0 ///
3	6	15	5.25 ... mai
4	10	18	4.2 ... mai
5	15	20	3.5 ... mai

In definitiva, conviene lanciare comunque due volte e proseguire con un terzo, ultimo, lancio solo se, con i primi due, non si sono raggiunti i 7$.

Per il calcolo del guadagno medio realizzabile con questa strategia, si consideri la situazione dopo i primi due lanci; si avra':

s =	con probabilita' p_s =
# 3	2/36
# 4	2/36
# 5	4/36
# 6	4/36
7	6/36
8	4/36
9	4/36
10	2/36
11	2/36

Nei casi # si procede ad un terzo lancio e li' il guadagno medio e' dato dalla (*) con $n=2$ (lanci eseguiti) ovvero $(1/6)[3s+21]$. Il valore richiesto e' dunque:

$$GM = \sum_{s=3}^{6} \frac{3s+21}{6} p_s + \sum_{s=7}^{11} s \cdot p_s = \frac{426}{216} + \frac{152}{36} = 6.19444$$

OSSERVAZIONE Operando 3 lanci, si hanno 216 casi possibili equiprobabili; il seguente programma calcola GM ''senza andare troppo per il sottile'' (la terna (X,Y,Z) viene fatta variare generando tutti i casi possibili)

```
10 FOR X=1 TO 6:FOR Y=1 TO 6:FOR Z=1 TO 6
20 IF X=Y THEN 60
30 A=X+Y:IF A>=7 THEN S=S+A:GO TO 60
40 IF Z=X OR Z=Y THEN 60
50 S=S+A+Z
60 NEXT Z:NEXT Y:NEXT X:PRINT''GM='';S/216
```

SOLUZIONE (b)

Sia ''n'' il numero di lanci dichiarato prima dell'inizio del gioco $(1 \leqslant n \leqslant 6)$; $A = \{1,2,3,4,5,6\}$
e
$B = \{\bar{x} = (x_1, x_2, .., x_n)$ con $x_i \epsilon A$ e $x_i \neq x_j$ per $i \neq j\}$.

L'insieme B e' formato da $D_{6,n}$ elementi (vedi 4.1) e il guadagno medio sara':

$$Gn = \frac{1}{6^n} \sum_{\bar{x} \epsilon B} \left(\sum_{k=1}^{n} x_k \right)$$

Il numero complessivo degli addendi nella ''doppia sommatoria'', ovvero il numero totale degli x_k, e' $n \cdot D_{6,n}$ e ogni elemento di A sara' presente un ugual numero di volte $N = n \cdot D_{6,n}/6$. Cio' significa che:

$$\sum_{\bar{x} \epsilon B} \left(\sum_{k=1}^{n} x_k \right) = 1 \cdot N + 2 \cdot N + \ldots + 6 \cdot N = 21 \cdot N$$

da cui:

$$Gn = \frac{1}{6^n} \cdot \frac{21 n \cdot D_{6,n}}{6} = \frac{2520\ n}{6^n \cdot (6 - n)!}$$

[essendo $D_{6,n}$ = 6!/(6-n)!]

Calcolando al variare di n si ha:

n	Gn
1	3.50
2	5.83
3	5.83
4	3.88
5	1.62
6	0.32

Conviene dunque dichiarare 2 oppure 3 lanci, guadagnando mediamente 5.83$ (per scegliere tra 2 e 3, occorrerebbe calcolare la varianza).

Come era prevedibile, il guadagno medio nella situazione (a) e' maggiore di quello realizzabile nella situazione (b) (sempre adottando le strategie ottimali).

LE SIMULAZIONI

Caso (a): **"I TRE AMICI"**

Il seguente programma di simulazione si propone due scopi:
- valutare il guadagno medio effettivo realizzato da un giocatore (Bruno) che adotti la strategia ottimale (valore teorico: 6.19444.$)
- confrontare i guadagni di Bruno con quelli ottenuti da Aldo e Carlo che adottano strategie non ottimali.

Ecco come si comportano i tre partecipanti:

ALDO. E' un tipo molto prudente. Se al primo lancio ottiene 6, si ferma (non volendo rischiare di perdere i 6 dollari); in caso contrario procede con un secondo, ultimo lancio.

BRUNO. Fa uso della strategia ottimale descritta.

CARLO. E' amante del rischio. Lancia comunque fino a tre volte; se dopo i tre tiri ha accumulato solo 6 dollari (il minimo possibile), opta per un quarto, ultimo lancio.

Il programma richiede il numero di simulazioni che si intendono effettuare e fornisce, per ogni giocatore:

(1) il guadagno medio per partita - GM (X PAR) -
(2) il relativo scostamento quadratico medio - DEV. STAN. -
(3) il numero di partite risultate ''utili'', ovvero quante si sono concluse con un qualche guadagno - PAR. UTILI -
(4) il ''guadagno medio per partita utile'', ovvero il rapporto tra il totale guadagnato e il valore (3) -GM(XP.U.)-

Principali variabili:
Per I=1,2,3 (rispettivamente Aldo, Bruno, Carlo); e':

G(I) : guadagno ottenuto nella partita in atto
U(I) : contatore partite utili
T(I) : somma dei dollari accumulati [ΣG(I)]
M(I) : somma dei quadrati dei guadagni [ΣG(I)*G(I)]
N$(I) : nome del giocatore
inoltre:
X,Y,Z,T : 1',2',3',4' uscita del dado
N : numero delle simulazioni
F : Flag: se F=1 le prime 3 uscite del dado sono tutte differenti, in caso contrario e' F=0.

Listato del programma I TRE AMICI

```
  5 DIM G(3),U(3),M(3),T(3),N$(3)
  7 N$(1)=''ALDO'':N$(2)=''BRUNO'':N$(3)=''CARLO''
 10 U=1:S=6:Q=4:A=0:D=2:H=3:INPUT''QUANTE PARTITE'';N
 30 FOR I=U TO N
 40 X=INT(RND(1)*S)+U:Y=INT(RND(U)*S)+U
 50 Z=INT(RND(1)*S)+U:T=INT(RND(U)*S)+U
 60 FOR K=U TO H:G(K)=A:NEXT K:F=A
 70 IF X=S THEN G(U)=S
 80 IF X=Y THEN 140
 90 W=X+Y:G(U)=-W*(X<S)-S*(X=S)
100 IF Z<>X AND Z<>Y THEN F=U
110 L=W+Z:G(D)=-W*(W>S)-L*(W<=S AND F=U)
120 IF F=A THEN 140
130 G(H)=-L*(L>S)-(L+T)*(T<>X AND T<>Y AND T<>Z)*
    (L<=S)
140 FOR K=U TO H:T(K)=T(K)+G(K):U(K)=U(K)-(G(K)>0):
    M(K)=M(K)+G(K)*G(K):NEXT K
160 NEXT I
220 FOR K=1 TO 3
230 PRINT N$(K):ME=T(K)/N:DS=SQR(M(K)/N-ME*ME)
235 GU=0:IF U(K)>0 THEN GU=T(K)/U(K)
240 PRINT ME,DS,U(K),GU:NEXT K
```

Commenti al listato

Istr. 5-10 : Introduzione (dimensionamenti ecc.)
Istr. 30-160 : Loop principale :
  40-60 lancio dei dadi; azzeramento guadagni e flag
  70-90 calcolo del guadagno di Aldo
  100-110 calcolo del guadagno di Bruno
  120-130 calcolo del guadagno di Carlo
  140 aggiornamento T(..), U(..), M(..)
Istr. 220-240: stampa risultati

Esempio di output del programma I TRE AMICI

```
NUM. PARTITE 1000

** ALDO
GM (X PAR)    5.498
DEV. STAN.    2.89758451
PAR.UTILI     852
GM (X P.U.)   6.45305165

** BRUNO
GM (X PAR)    6.067
DEV. STAN.    4.08907215
PAR.UTILI     708
GM (X P.U.)   8.56920904

** CARLO
GM (X PAR)    5.478
DEV. STAN.    5.92735321
PAR.UTILI     520
GM (X P.U.)   10.5346154
```

Un confronto tra la statistica di Aldo e quella di Carlo mostra come la strategia prudente del primo lo porti a guadagnare spesso (ben 852 partite utili su 1000) ma poco (6.45$ per ogni partita utile), al contrario di quanto avviene per Carlo che vince raramente (solo 520 partite utili), ma quando vince, vince bene (10.53$ per partita utile).

I valori GM(XP.U.) si comprendono meglio se si nota che i minimi e i massimi guadagni in una partita utile sono:

per Aldo	min	=	3 $ (es. uscite 1-2)
	max	=	11 $ (es. uscite 5-6)
per Bruno	min	=	6 $ (es. uscite 1-2-3)
	max	=	12 $ (es. uscite 4-2-6)
per Carlo	min	=	7 $ (es. uscite 1-2-4)
	max	=	15 $ (es. uscite 4-5-6)

Naturalmente, il valore piu' interessante e' il primo — GM (X PAR) —,che, presumibilmente, risultera' maggiore per Bruno.

Caso (b): **"I SEI AMICI"**

In modo analogo a quanto visto per il caso (a), si vogliono confrontare i risultati ottenuti da persone che si comportano diversamente. In questo caso simuleremo tutte le strategie possibili: prima di iniziare il gioco, Aldo dichiara un lancio, Bruno ne dichiara 2, Carlo 3, Dino 4, Ettore 5, Flavio 6.

Dopo le assegnazioni e i dimensionamenti di rito, il programma richiede il numero di partite (istr. 20). Le istruzioni dalla 30 alla 200 costituiscono il ciclo principale; per ogni gara il dado viene lanciato un certo numero di volte:
istr. 30 X(1) = uscita del 1' lancio
istr. 60 T = uscita di un successivo lancio (2',3',..)
istr. 80 si controlla se T e' diverso dalle precedenti uscite...
istr. 90 ...in caso negativo, viene posto X(K) = T e si ritorna alla 60 per un nuovo lancio.

L'istruzione 150 viene eseguita non appena T fornisce un valore identico a uno dei lanci precedenti (naturalmente non si possono avere piu' di 6 uscite diverse); la variabile W contiene allora il massimo indice dei giocatori che realizzano un guadagno positivo. Per i giocatori 1,2,..,W viene valutato il guadagno G(L) e ne viene aggiornata la situazione (T(L) = totale accumulato, U(L) = numero partite utili e Q(L) = somma dei quadrati dei guadagni).

All'istruzione 180 i guadagni vengono azzerati e si procede ad una nuova gara.

Esempio:
1' lancio 4: X = (4,....)
2' lancio T = 2: e' T ≠ X(1) allora X = (4,2,....)
3' lancio T = 6: e' T ≠ X(K) K = 1,2 quindi X = (4,2,6,....)
4' lancio T = 2: e' T = X(2): salto alla 150 con W = 3.

Vincono Aldo, Bruno, Carlo e guadagnano:
Aldo 4 $: G(1)
Bruno 6 $: G(2) = G(1) + X(2)
Carlo 12 $: G(3) = G(2) + X(3)
(Nel ciclo 150-170 viene utilizzato l'elemento G(0) = 0).
Per gli altri giocatori la situazione rimane immutata.

Le istruzioni 400-420 realizzano la stampa dei risultati; con la successiva pressione di un tasto qualsiasi, vengono visualizzati i valori relativi a Bruno. poi Carlo ecc.

Listato del programma I SEI AMICI

```
 10 DIM X(6),G(6),T(6),U(6),Q(6),N$(6)
 11 FOR F=1 TO 6:READN$(F):NEXT F
 12 DATA ALDO,BRUNO,CARLO,DINO,ETTORE,FLAVIO
 20 U=1:E=6:D=2:Z=0:INPUT''QUANTE PARTITE'';N
 30 FOR S=U TO N:X(U)=INT(RND(U)*E)+U
 60 FOR K=D TO S:T=INT(RND(U)*E)+U:W=K-U
 80 FOR H=U TO W:IF T=X(H) THEN 150
 90 NEXT H:X(K)=T:NEXT K:W=K-U
150 FOR L=U TO W:G(L)=G(L-U)+X(L):T(L)=T(L)+G(L)
170 U(L)=U(L)+U:Q(L)=Q(L)+G(L)*G(L):NEXT L
180 FOR M=U TO E:G(M)=Z:NEXT M
200 NEXT S
400 FOR F=1 TO 6:PRINT F,N$(F):ME=T(F)/N:DS=
    SQR(Q(F)/N-ME*ME)
415 GU=0:IF U(F)>0 THEN GU=T(F)/U(F)
420 PRINT ME,DS,U(F),GU:NEXT F
```

Esempio di output del programma I SEI AMICI

```
QUANTE PARTITE   1000

1 ** ALDO
GM (X PAR)          3.526
DEV. STAN.          1.68858639
PAR. UTILI          1000
GM (X P.U.)         3.526

2 ** BRUNO
GM (X PAR)          5.802
DEV. STAN.          3.21291083
PAR. UTILI          831
GM (X P.U.)         6.98194946

3 ** CARLO
GM (X PAR)          5.895
DEV. STAN.          5.43378092
PAR. UTILI          563
GM (X P.U.)         10.4706927

4 ** DINO
GM (X PAR)          3.94
DEV. STAN.          6.44394289
PAR. UTILI          278
GM (X P.U.)         14.1726619
```

5 ** ETTORE

GM (X PAR)	1.53
DEV. STAN.	4.98268001
PAR. UTILI	87
GM (X P.U.)	17.58620069

6 ** FLAVIO

GM (X PAR)	0.336
DEV. STAN.	2.63497704
PAR. UTILI	16
GM (X P.U.)	21

Il ''guadagno medio per partita'' mostra, come ci si aspettava, che Bruno e Carlo adottano le strategie migliori.

E' possibile confrontare il numero di ''partite utili'' ottenute con il relativo valore atteso: per il giocatore di indice ''k'' ($k=1$ Aldo, $k=2$ Bruno,...), la probabilita' che una partita si concluda con un guadagno positivo e' $p_k = D_{6,k}/6^k$, per cui le partite utili saranno circa $N \cdot p_k$ ($N=$ numero simulazioni).

Per $N=1000$ si ha:

indice giocatore	valore atteso per P.U. (circa)
1	1000 (ovvio)
2	833
3	556
4	278
5	93
6	15

valori questi che ben si accordano con quelli simulati.

2.14 AL RIALZO

Un programma associato:

''SFIDA''

Un giocatore lancia un dado regolare una o piu' volte. Fino a che il punteggio ottenuto e' maggiore o uguale a quello dei lanci precedenti, egli riceve tanti dollari quanti sono i punti riportati,

che si sommano al totale precedentemente guadagnato. Se ottiene un punteggio inferiore, il gioco ha termine e il giocatore perde tutto quanto aveva accumulato.

Egli puo' decidere di fermarsi quando lo desidera.

Come deve comportarsi per massimizzare la sua speranza di guadagno?

Esempio 1:

esce ''1'': decide per un nuovo lancio
..esce ''4'': decide per un nuovo lancio
..esce ''4'': decide di fermarsi.
Il giocatore ha guadagnato 9 $

Esempio 2:

esce ''2'': decide per un nuovo lancio
..esce ''1'': Il gioco ha termine con guadagno nullo.

Il programma associato a questo problema consente all'utilizzatore di definire la propria strategia e di misurarsi contro il calcolatore.

La strategia ottimale, adottata dal calcolatore, verra' descritta nel paragrafo ''soluzione'', paragrafo che andra' letto dopo aver giocato, magari piu' volte, con il programma SFIDA (si eviti di esaminarne il listato).

Strategie

Nel problema proposto, una strategia e' una ''funzione di decisione'' che, in base alle informazioni di cui disponiamo al momento, ci dice cosa dobbiamo fare: fermarci o continuare?

Un esempio di strategia potrebbe essere:
1. Lancio il dado la prima volta
2. se oggi e' lunedi', lancio il dado una seconda volta e mi fermo; se non e' lunedi', mi fermo subito accettando l'uscita del primo lancio.

E' facile rendersi conto che (se ci si propone di guadagnare il piu' possibile) quella descritta e' una strategia insensata.

Da tutte le strategie possibili isoliamo quelle piu' promettenti. Due sono le informazioni che devono contribuire a determinare il nostro comportamento (dopo il primo lancio):

- il valore ottenuto nell'ultimo lancio
- la somma che si guadagna se si opta per ''l'arresto''.

Confrontiamo due esempi:

Es.1 A: 2/2/2/2/2/
B: 1/1/1/1/6/

Nei casi A e B sono stati eseguiti 5 lanci; in entrambi, se ci si

ferma, si guadagnano 10 $. Nel caso B, proseguire e' assai piu' rischioso che nel caso A: il risultato dell'ultimo lancio e' importante.

Es.2 A: 1/1/1/1/3/
B: 3/3/3/3/3/

Sia in A che in B, i lanci sono 5; l'ultima uscita e' 3. Fermandosi si guadagnano 7 $ o 15 $ rispettivamente. Proseguendo nel caso B, si mette in gioco una somma maggiore che nel caso A: il ''totale accumulato'' e' importante.
(Non interessa invece il numero di lanci eseguiti; se A: 3/3/ B: 1/2/3/, i giocatori A e B si trovano in situazione identica).

Una strategia che si rispetti sara' cosi' strutturata:

1. Eseguo il primo lancio
2. Se x e' l'uscita dell'ultimo lancio eseguito e S e' la somma che riceverei scegliendo di fermarmi, decidero' di

PROSEGUIRE se $S < f(x)$ (e rientro al punto 2)
FERMARMI se $S \geqslant f(x)$.

Si tratta allora di scegliere i valori ''di soglia'' associati ad ogni uscita.

Avremo f(1) = ''infinito'': se l'ultimo lancio ha dato 1, conviene senz'altro proseguire.

Inoltre sara' $f(2) \geqslant f(3) \geqslant \ldots \geqslant f(6)$: al crescere di x aumenta il rischio di perdere tutto e conviene accontentarsi di una somma minore.

Un esempio di strategia ''sensata'' potrebbe essere:

x	f(x)
1	infinito
2	30
3	20
4	10
5	5
6	2

Nel programma SFIDA, l'utilizzatore e' invitato ad impostare la propria strategia: per $x = 2,3,\ldots,6$ si inserira' il valore f(x) che si ritiene piu' conveniente. Il significato e' ormai chiaro: porre f(2) = 30 equivale a dire ''se all'ultimo lancio ho ottenuto 2 e non ho ancora raggiunto i 30 $, intendo proseguire nel gioco''.

Se, per qualche x, inserite un valore $f(x) < x$, intendete ''fermar-

vi'' non appena il dado fornisce un'uscita pari a x. In questo caso, il valore effettivo di f(x) non ha importanza (l'effetto di f(6) = 2 e f(6) = 3 e' identico: non appena esce il 6 e' S $\geqslant$ 6 ed in entrambi i casi si opta per l'arresto).

I valori inseriti vengono caricati nelle componenti da 2 a 6 del vettore G(I) (istr. 40) precedentemente dimensionato (istr. 10) (G(0) e G(1) non vengono utilizzate). In modo analogo vengono caricati in C(I) i valori di soglia della strategia adottata dal calcolatore (istr. 80).

Dopo aver impostato il numero di partite che si intendono simulare (NS nella 100), ha inizio la sfida vera e propria al termine della quale il programma fornisce (istr. 600):

- 1 Il numero di dollari totalizzato dal giocatore
- 2 Il relativo guadagno medio per gara
- 3 Il numero di dollari totalizzato dal calcolatore
- 4 Il relativo guadagno medio per gara

Poiche' il calcolatore adotta la strategia migliore, c'e' da aspettarsi che vinca la sfida (se vi batte ''di misura'', avete compreso perfettamente il meccanismo e sapete ben intuire le probabilita' degli eventi in gioco; se il calcolatore ''stravince'', evitate i giochi d'azzardo; se siete voi a vincere, siete bravi ... e fortunati).

Le variabili utilizzate nella simulazione sono:

- U: valore fornito dal dado nell'ultimo lancio
- G,C: flags. Se G = 0, il giocatore intende ''proseguire''; se G = 1, il giocatore intende ''fermarsi''. ''C'' e' l'analogo valore per il calcolatore
- SG,SC: SG e' la somma accumulata dal giocatore nella singola gara ovvero cio' che riceve se decide di fermarsi (cosi' SC per il calcolatore)
- TG,TC: totale dollari nelle varie partite per giocatore e calcolatore rispettivamente
- X: risultato del ''nuovo'' lancio del dado
- W: contatore delle simulazioni.

La simulazione viene realizzata nel ciclo 120-540:

200: viene eseguita una sola volta per ogni gara. U rappresenta qui l'uscita del primo lancio; le somme accumulate vengono poste uguali ad U; C e G azzerati.
240-280: si controlla se il giocatore intende proseguire, in caso negativo viene aggiornato il valore TG.
300-340: come per 240-280 in relazione al calcolatore.
360: se Giocatore e Computer sono entrambi ''fermi'', si procede ad una nuova gara.

380: determina la nuova uscita del dado (X)
400-440: se il giocatore ha deciso di proseguire e $X \geqslant U$, SG viene opportunamente incrementato; se $X < U$, la somma accumulata viene azzerata e il giocatore ''si ferma'' (SG=0:G=1).
460-500: come per 400-440 in relazione al calcolatore
520: U viene posto uguale a X; la gara prosegue (istr. 240).
540: chiude il ciclo aperto alla 120.

Il listato del programma SFIDA e' riportato dopo la soluzione.

Esempio di utlizzo del programma SFIDA

```
SCEGLI I VALORI DI SOGLIA
PER LA TUA STRATEGIA
*  2 *
? 30
*  3 *
? 20
*  4 *
? 10
*  5 *
?  5
*  6 *
?  2
NUM. SIMULAZ.? 500
   $ GIOC.        2746
G.M. GIOC.        5.492
   $ COMP.        2796
G.M. COMP.        5.592
```

SOLUZIONE

Supponiamo di aver effettuato con successo il lancio di indice ''n'' e di dover decidere se proccedere con l'n+1-esimo. Sia:

''s'' la somma accumulata con i primi n tiri

''u'' il valore ottenuto nell'ultimo lancio (n-esimo).

RINUNCIANDO si guadagna ''s''

PROSEGUENDO si guadagna: s+x se $x \geqslant u$
: 0 se $x < u$

dove x=valore uscito all'n+1-esimo lancio.

Nell'ultimo caso le uscite utili sono: u,u+1,...,6 che sono in numero di 7-u. Poiche' ognuno dei valori ha probabilita' 1/6, decidendo di continuare si guadagnera' complessivamente:

s+u con probabilita' 1/6

s+u+1 con probabilita' 1/6

...   ...
s+6   con probabilita' 1/6

Il guadagno medio, per un giocatore che decida di proseguire, e' dunque:

$$G=\frac{1}{6}(s+u)+\frac{1}{6}(s+u+1)+\ldots+\frac{1}{6}(s+6)=\frac{1}{6}\left[(7\text{-}u)s+\sum_{i=u}^{6} i\right]$$

E':

$$\sum_{i=u}^{6} i = \sum_{i=1}^{6} i - \sum_{i=1}^{u-1} i = \frac{6\cdot 7}{2} - \frac{u(u-1)}{2} = \frac{42 - u^2 + u}{2}$$

per cui

$$G = \frac{1}{6} \cdot \frac{14s - 2us + 42 - u^2 + u}{2}$$

Converra' continuare se G>s, ovvero se

$$14s - 2us + 42 - u^2 + u > 12s$$
$$2s(1-u) > u^2 - u - 42$$
$$2s(u-1) < (u+6)(7-u)$$

Per u=1 la diseguaglianza e' vera per ogni s ed e' ovvio che conviene continuare. Per u>1 conviene proseguire se:

$$s < \frac{(u+6)(7-u)}{2(u-1)}$$

Calcolando l'espressione per u=2,3,...,6 si ottiene il seguente "piano di gioco" ovvero i valori ideali per f(u):

se l'ultimo lancio ha dato...	**... conviene proseguire se "s" e' minore di:**
1	proseguire comunque
2	20
3	9
4	5
(*) 5	2.75
(*) 6	1.2

Nei casi (*), si ha s $\geqslant$ 5: non appena si ottiene 5 o 6 conviene sempre fermarsi.

Listato del programma SFIDA

```
 10 DIM G(6),C(6)
 40 FOR I=2 TO 6:INPUT G(I):NEXT I
 80 FOR I=2 TO 6:READ C(I):NEXT I:DATA 20,9,5,2.75,1.2
100 INPUT''QUANTE SIMUL.'';NS
120 FOR W=1 TO NS
200 U=INT(RND(1)*6)+1:G=0:C=0:SG=U:SC=U
240 IF G=1 THEN 300
260 IF U=1 OR SG<G(U) THEN G=0:GO TO 300
280 G=1:TG=TG+SG
300 IF C=1 THEN 360
320 IF U=1 OR SC<C(U) THEN C=0:GO TO 360
340 C=1:TC=TC+SC
360 IF G*C>0 THEN 540
380 X=INT(RND(1)*6)+1
400 IF G=1 THEN 460
420 IF X>=U THEN SG=SG+X:GO TO 460
440 SG=0:G=1
460 IF C=1 THEN 520
480 IF X>=U THEN SC=SC+X:GO TO 520
500 SC=0:C=1
520 U=X:GO TO 240
540 NEXT W
600 PRINT TG,TG/NS,TC,TC/NS
```

2.15 DADOMAX

Un programma associato:

"DADOMAX"

Un giocatore dispone di un dado regolare a sei facce; dopo aver dichiarato quanti lanci intende effettuare (n), procedera' con gli n lanci e ricevera' un numero di dollari pari al massimo dei punteggi ottenuti.

Per giocare, deve prima pagare una tassa proporzionale al numero di lanci dichiarato: $T = a \cdot n$.

Esempio: sia a = 0.1 \$, decido di lanciare 4 volte e quindi pago 0.4\$.

Lancio: ottengo 2/4/2/3; ricevo 4\$: ho guadagnato 3.6\$.

Determinare, fissato ''a'' (''costo di un lancio''), il valore n' piu' conveniente.

SOLUZIONE

Per ogni n fissato, sia x_i il punteggio ottenuto all'i-esimo lancio e $x'_n = \max_{1 \leqslant i \leqslant n} x_i$.

Vogliamo determinare la legge della variabile aleatoria x'_n, ovvero $P[x'_n = k]$ per $k = 1,2,..,6$.

Fissati ''n'' e ''k'', la quantita' $P[x'_n = k]$ puo' essere calcolata come rapporto ''casi favorevoli/casi possibili''.

I ''casi possibili'' sono tanti quanti le disposizioni con ripetizione di 6 oggetti presi a gruppi di ''n'':

$D'_{6,n} = 6^n$

(vedi 4.2).

I ''casi favorevoli'' sono costituiti da quelle disposizioni con ripetizione che contengono ''k'' come massima uscita: queste sono le disposizioni degli elementi $\leqslant k$ eccettuate quelle che contengono solo elementi $\leqslant k\text{-}1$. I casi favorevoli sono dunque in numero di

$$D'_{k,n} - D'_{k-1,n} = k^n - (k-1)^n.$$

Allora

$$p_k = P[x'_n = k] = \frac{k^n - (k-1)^n}{6^n}.$$

La variabile aleatoria x'_n assume quindi il valore k con probabilita' p_k con $1 \leqslant k \leqslant 6$. Il suo valore medio e':

$$E[x'_n] = \sum_{k=1}^{6} k \cdot \frac{k^n - (k-1)^n}{6^n} = 6 - \sum_{k=1}^{5} \left(\frac{k}{6}\right)^n$$

[Il risultato si ottiene sviluppando la sommatoria e procedendo ad opportuni raccoglimenti, oppure, piu' rapidamente, notando che
$E[x'_n] = P(x'_n > 0) + P(x'_n > 1) + \ldots + P(x'_n > 5)$
con $P(x'_n \leqslant k) = (k/6)^n$ per cui $P(x'_n > k) = 1 - (k/6)^n$.]

$E[x'_n]$ rappresenta dunque il ricavo atteso per un giocatore che decida di lanciare ''n'' volte (come ci si aspettava, il valore si avvicina a 6 al crescere di n).

Il guadagno medio e':

$$G(n) = 6 - [\sum_{k=1}^{5} \left(\frac{k}{6}\right)^n] - a \cdot n$$

e il problema consiste nel determinare $n' \geqslant 1$ che rende massimo G(n).

$G(1) = 3.5 - a > 0$ per $a < 3.5$: se la tassa e' maggiore o uguale a 3.5 \$, non conviene giocare.

Per ogni a fissato minore di 3.5, e' $G(n) \leqslant 6 - a \cdot n$ e il gioco perde di convenienza per $n > 6/a$ (risultato prevedibile dato che il ricavo non supera comunque i 6 \$).

Un modo per determinare n' puo' essere quello di calcolare G(k) per k intero da 1 a INT(6/a) e scegliere il valore k che rende massima G(..).

Si puo' fare di meglio: consideriamo la differenza

$$G(n+1) - G(n) = \left[\sum_{k=1}^{5} \left(\frac{k}{6}\right)^n \left(1 - \frac{k}{6}\right) \right] - a = D_n - a$$

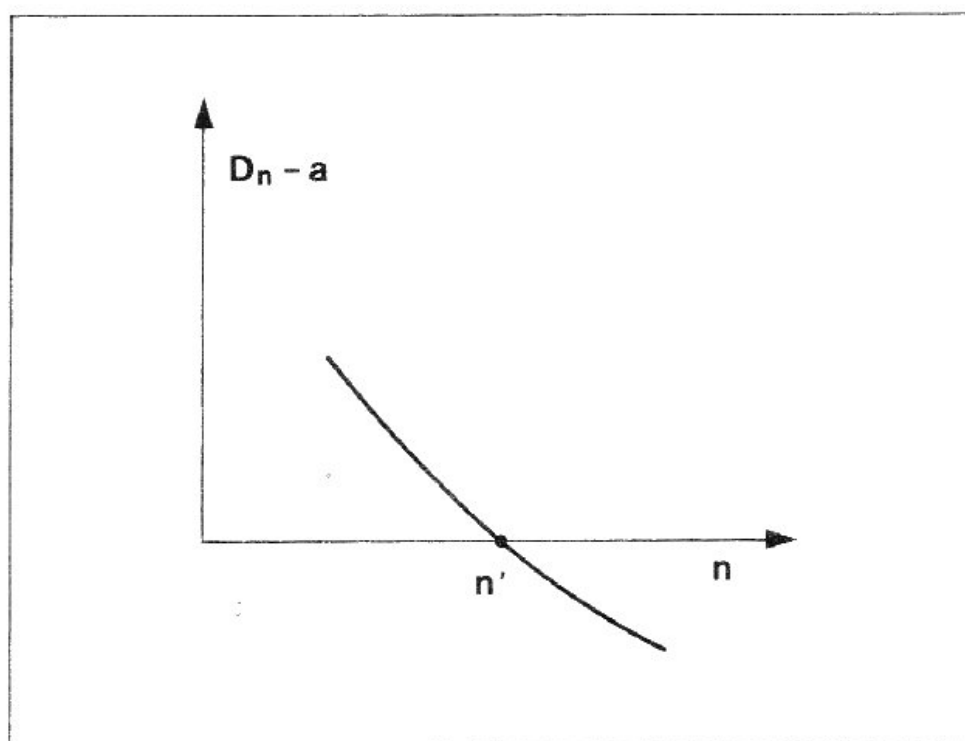
Poiche' $(k/6) < 1$ e' $(k/6)^{n+1} < (k/6)^n$ e anche $D_{n+1} - a < D_n - a$ (vedi fig. 14) quindi:

per $n < n'$ e' $G(n+1) > G(n)$

e per ogni $n > n'$ e' $G(n+1) < G(n)$.

Questo significa che e' possibile determinare la soluzione calcolando G(k) k = 1,2,... fermandosi non appena G(..) inizia a diminuire.

Il programma DADOMAX determina, in questo modo, il numero ideale di lanci in relazione alla tassa A fissata.

Fig. 14 *Grafico di G(n + 1)-G(n) (problema 2.15)*

Listato del programma DADOMAX

```
10 INPUT''TASSA PER LANCIO'';A:B=INT(6/A):U=1:C=5:Z=0:Q=6
30 FOR I=U TO B:S=Z
50 FOR K=U TO C:S=S+(K/Q)↑I:NEXT K
60 G=Q-S-A*I
70 IF G>M THEN M=G:N=I:GO TO 80
75 GO TO 90
80 NEXT I
90 PRINT N,M
```

Esempi

a	n'	G(n')
0.00001	54	5.999
0.0001	41	5.995
0.001	29	5.966
0.01	16	5.784
0.05	9	5.328
0.1	6	4.960
0.5	2	3.472
1	1	2.500

2.16 IL LINGUAGGIO

Un programma associato:
''LINGUAGGIO''

PREMESSA

Il programma che segue costituisce un esempio di applicazione dei PROCESSI DI MARKOV al calcolatore.

Per la comprensione del programma, non e' necessaria una conoscenza completa dei processi di Markov; bastera' pensare ad un ''sistema'' che puo' trovarsi in ''n'' stati distinti e che puo' passare da uno stato all'altro (o rimanere nello stato attuale) con certe probabilita'.

Processi di Markov

Quando la probabilita' di passaggio dallo stato ''i'' (presente) allo stato ''j'' (futuro) dipende esclusivamente dalla situazione presente e non da tutta ''la storia precedente'' (ovvero da ''come vi si e' giunti''), si ha un processo di Markov.

Le probabilita' di passaggio da uno stato all'altro vengono dette probabilita' di transizione. Se queste probabilita' non variano

al passare del tempo, il processo si dice stazionario.

I valori delle probabilita' di transizione vengono agevolmente rappresentati in una matrice di transizione in cui l'elemento p_{ij} rappresenta la probabilita' che il sistema, attualmente allo stato "i", passi, all'istante successivo, allo stato "j" (il tempo viene qui considerato come grandezza discreta).

Ad esempio, consideriamo un sistema con 3 possibili stati: A, B e C.

La matrice di transizione potrebbe essere:

	a	A	B	C	al tempo t+1
da...	A	1/2	1/4	1/4	
al tempo t	B	0	0	1	
	C	1/2	0	1/2	

questo significa che, se il sistema si trova nello stato A, passera', all'istante successivo, allo stato B con probabilita' 0.25; allo stato C con probabilita' 0.25, mentre con probabilita' 0.5 rimarra' nello stato A.

La seconda riga ci dice che, ogniqualvolta il sistema si trova nello stato B, passa obbligatoriamente allo stato C; la terza ci informa che dallo stato C si passa, con la stessa probabilita', allo stato A o si rimane nello stato C.

I valori della matrice, in quanto probabilita', sono compresi tra 0 e 1 e la somma dei valori di ogni riga e' uguale a 1.

Se il processo e' stazionario, una possibile successione di stati del sistema proposto nell'esempio potrebbe essere AABCACC-CAAAAC..., mentre la successione ABACB... risulta inammissibile in quanto dallo stato B non si puo' passare allo stato A ne' da C a B.

IL LINGUAGGIO

La successione delle lettere alfabetiche che formano le parole e' strutturata in modo diverso da lingua a lingua.

Nelle parole italiane, le coppie di lettere come "ma", "ca", "le" sono senz'altro piu' frequenti della coppia "ps" (presente, ad esempio, in termini come "lapsus" o "psicanalisi"); nella lingua inglese sono invece abbastanza frequenti le coppie di lettere come "wh" (in "when", "who", "why" ecc.), inesistenti nella lingua italiana.

Consideriamo ora un sistema che puo' assumere 27 stati distinti: le 26 lettere dell'alfabeto + lo "spazio" inteso come carattere di separazione tra una parola e l'altra.

Una frase sara' intesa come una serie di passaggi da uno stato

all'altro, ad esempio nella frase "IO CORRO" il sistema passa dallo stato iniziale "I" allo stato "O" quindi allo stato "SPAZIO", poi allo stato "C" ecc.

Nella prima parte del programma in esame, il calcolatore accetta in entrata un testo qualsiasi fino a quando non viene immesso il simbolo "*" che segnala la fine del testo. In questa fase, simboli diversi dalle lettere alfabetiche vengono stampati sullo schermo, ma considerati come "spazio".

Man mano che l'immissione del testo procede, il calcolatore aggiorna i valori di una matrice quadrata T di 27x27 elementi: agli indici di linea e colonna corrispondono le lettere alfabetiche piu' lo "spazio" e ogni elemento della matrice rappresenta il numero di volte che una determinata lettera (indice I) e' stata immediatamente seguita da un'altra determinata lettera (indice J).

Ad esempio, supponendo l'alfabeto costituito dalle sole tre lettere A, B e C, l'immissione della stringa ABBACCCCACC* originerebbe la matrice:

'T'	A	B	C
A	0	1	2
B	1	1	0
C	1	0	4

Una volta costruita, la matrice T viene trasformata in una matrice di transizione dividendo ogni elemento per la somma dei valori della propria linea (se i valori di una certa linea sono tutti nulli, questa operazione non viene eseguita per detta linea).

Proseguendo nell'esempio precedente e indicando con S(i) la somma dei valori dell'i-esima linea, si ha S(1) = 3; S(2) = 2; S(3) = 5, per cui la matrice viene trasformata nella seguente:

	A	B	C
A	0	1/3	2/3
B	1/2	1/2	0
C	1/5	0	4/5

A questo punto il calcolatore stampa un testo casuale in cui la probabilita' di successione di una data coppia di lettere viene fornita dalla matrice T. L'output del testo prosegue in modo indefinito; premendo un tasto qualunque si ottiene la sospensione del-

l'elaborazione che viene ripresa premendo ulteriormente un tasto.

Se il testo in entrata e' sufficientemente esteso (diciamo di almeno 60 parole), quello in uscita risulta essere una buona simulazione del liguaggio proposto: se inseriamo una pagina in lingua italiana, otteniamo un testo casuale che ''sa di italiano'', cosi' per il francese, il tedesco ecc.

Poiche' lo ''spazio'' e' trattato alla stregua di qualsiasi altra lettera, e' possibile ottenere in uscita un testo con qualche parola di lunghezza eccessiva; inoltre, se in entrata sono presenti parole con lettere doppie, es. ''io corro'', in uscita e' possibile avere parole come ''corrro'' o ''corrrro''.

Nonostante queste limitazioni, il programma fornisce risultati interessanti che, naturalmente, sono tanto migliori quanto maggiormente risulta esteso il testo in entrata.

Listato del programma LINGUAGGIO

```
  7 DIM A(26,26)
  9 T$='' '':PRINT CHR$(17)+CHR$(29)+CHR$(29)+''-'';
 10 GET A$:IF A$='''' THEN 10
 20 IF A$=''*'' THEN 205
 30 PRINT A$;:J=ASC(A$)-64:I=ASC(T$)-64
 60 IF J<1 OR J>26 THEN J=0
 70 IF I<1 OR I>26 THEN I=0
 80 A(I,J)=A(I,J)+1:T$=A$:GO TO 10
205 FOR I=0 TO 26
210 S=0
220 FOR J=0 TO 26:S=S+A(I,J):NEXT J
240 IF S=0 THEN 280
250 FOR J=0 TO 26:A(I,J)=A(I,J)/S:NEXT J
280 NEXT I
300 L=0
310 X=RND(1):S=0:R$='''':S$=''''
320 FOR I=0 TO 26:S=S+A(L,I)
340 IF S>X THEN 360
350 NEXT I
360 IF I=0 OR I=27 THEN PRINT'' '';:GO TO 380
370 PRINT CHR$(I+64);
380 L=I
385 IF I>26 THEN L=26
410 GO TO 310
```

Commenti al listato

Istr. 10-30: accetta il testo, se terminato va alla 205

Istr. 60-70: caratteri diversi dalle lettere alfabetiche vengono considerati ''spazi''

Istr. 80: aggiornamento valori della matrice A

Istr. 205-280: trasforma A in matrice di transizione

Istr. 310-370: sceglie il carattere da stampare in funzione dei valori di A - Nel ciclo 320-350, la variabile S, inizialmente nulla, viene incrementata successivamente dai valori contenuti nella linea di indice L (ultima lettera stampata). Quando S supera X, l'indice I raggiunto definisce la nuova lettera da stampare cosicche' la probabilita' di stampa di una certa lettera e' quella definita dalla linea L della matrice A-.

Poiche' la matrice A viene sistematicamente aggiornata, la correzione di caratteri erroneamente introdotti risulterebbe impossibile. Nel programma registrato su disco, tuttavia, e' consentito cancellare e riscrivere i caratteri nell'ambito della stessa riga; questi vengono infatti caricati in una variabile di transito e la matrice A modificata solo a linea completa.

2.17 L'OPINIONE PUBBLICA

Due programmi associati:

''OPINIONE/V''
''OPINIONE/S''

Nella rubrica ''Ricreazioni al calcolatore'' apparsa sul numero di giugno 1985 di ''Le Scienze'', A.K.Dewdney proponeva la redazione di cinque semplici programmi legati ad altrettanti problemi. Il programma che segue costituisce un ampliamento di uno dei temi proposti (VOTERS e ANTIVOTERS) e, con un'opportuna interpretazione dei due ''partiti'', un medico potrebbe vedervi un modello semplificato di fenomeni a lui familiari.

Consideriamo una matrice di L linee e C colonne che identifichiamo con un rettangolo composto da TC=LxC ''celle'' (vedi fig.15).

Pensiamo ora di incollare il lato AB con il lato CD, arrotolando il rettangolo in modo da avere un cilindro e di unire le circonferenze γ 1 e γ 2 come mostrato in fig.16.

La superficie del solido che si ottiene, cio' che i matematici chiamano TORO (una specie di ciambella), e' ''priva di bordi''. Proprio questa superficie quadrettata sara' il teatro della nostra storia.

Inizialmente, il programma OPINIONE assegna casualmente a ciascuna delle celle un elettore che puo' essere un CONSERVATORE (simbolo ''*'') o un RIFORMISTA (simbolo ''.''); si comincia dunque da una situazione in cui i sostenitori dei due partiti sono ben mescolati sia come consistenza numerica sia come disposizione geografica.

Fig. 15 *Matrice di 4 linee e 7 colonne*

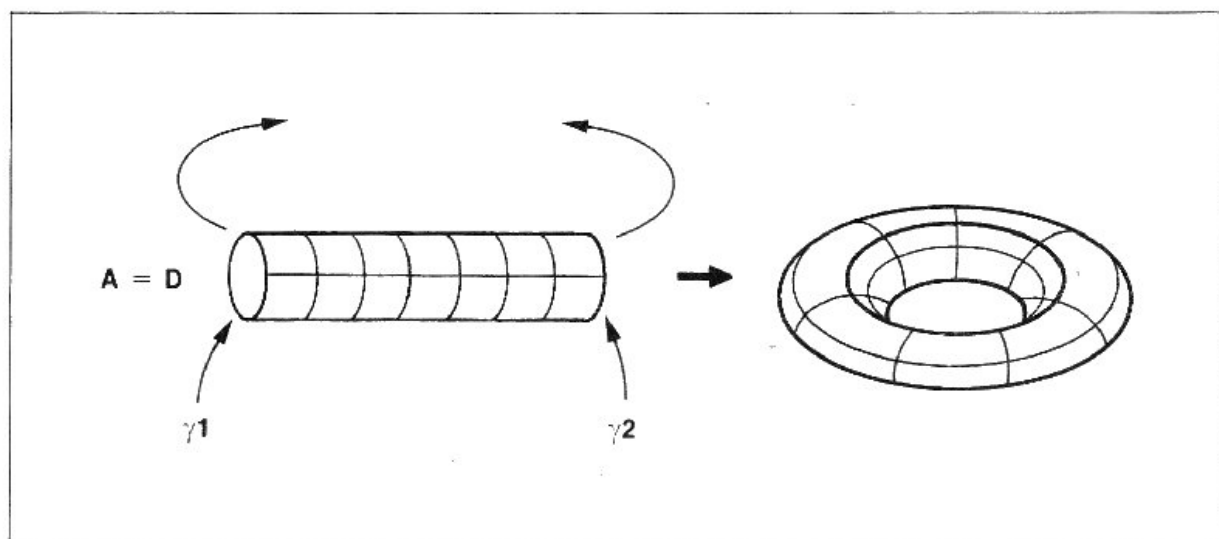
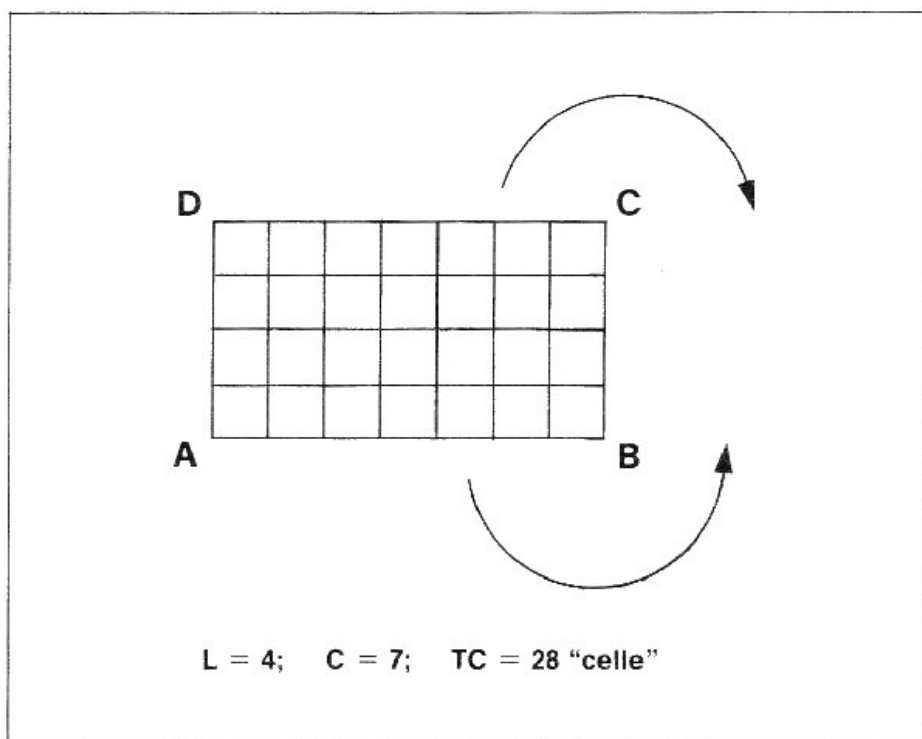
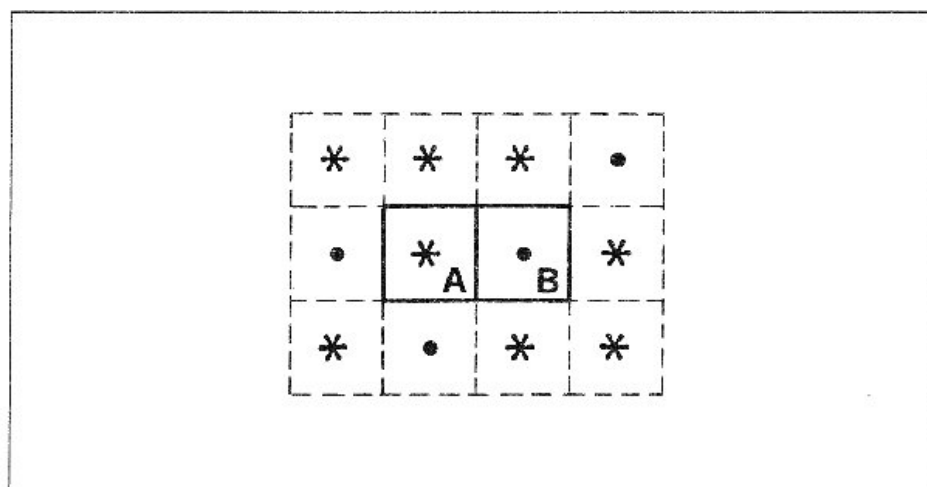



Fig. 16 *Il rettangolo viene trasformato in toro*

A questo punto si apre il ciclo principale. Una coppia di celle contigue viene scelta a caso (viene generata una coppia di coordinate, diciamo (2,5) e viene selezionata, sempre casualmente, una delle quattro celle adiacenti; nell'esempio, la (1,5), la (3,5), la (2,4) o la (2,6)).

Se i due elettori sono della stessa idea, non accade nulla e il programma rientra al punto precedente per la scelta di una nuova coppia. Quando vengono trovate opinioni differenti, la variabile "tempo" viene incrementata e ciascuno degli occupanti cerca di convincere l'altro a "cambiare bandiera".

I due elettori sono circondati da 10 vicini (parenti, amici, colle-

Fig. 17 *Coppia di elettori circondata da 10 "vicini"*

ghi ...) la cui opinione puo' influenzare l'opera di convincimento di uno o dell'altro (figure 17 e 18).

Nel programma OPINIONE, l'evento

{Il conservatore (*) convince il riformista (.) a passare dalla sua parte}

ha probabilita'

P[*] = (1-2p)x/10 + p

dove "x" e' il numero di conservatori che circonda la coppia (x = 7 in fig.17; x = 10 in fig.18) e "p" e' una probabilita' richiesta all'inizio del programma.

Il valore assegnato a "p" regola l'influenza di "x" nel computo di P[*] (fig.19):

- Se p = 0 e' P[*] = x/10. In questo caso l'esito della disputa e' fortemente influenzato dall'opinione del vicinato. Nel caso di fig.18, e' addirittura P[*] = 1 e l'elettore riformista viene certamente convinto a cambiare opinione (la pressione dell'opinione pubblica e' schiacciante).

- Quando 0 < p < 1/2, c'e', anche nel caso di fig.18, la possibilita' che sia l'elettore "in minoranza" a convincere l'avversario. La probabilita' che cio' si verifichi aumenta al crescere di "p".

- Se p = 1/2 e' P[*] = 1/2 qualunque sia x: l'opinione dei vicini non ha importanza. Con probabilita' del 50% il conservatore convincera' il riformista, con probabilita' del 50% succedera' il contrario.
  In questo caso il programma OPINIONE si comporta come il programma VOTERS proposto sulla rivista "Le Scienze".

- Per p > 1/2 assistiamo ad una inversione di tendenza. Il soggetto "piu' forte" si lascia convincere facilmente a cambiar partito e

Fig. 18 *Coppia di elettori circondata da 10 "vicini"*

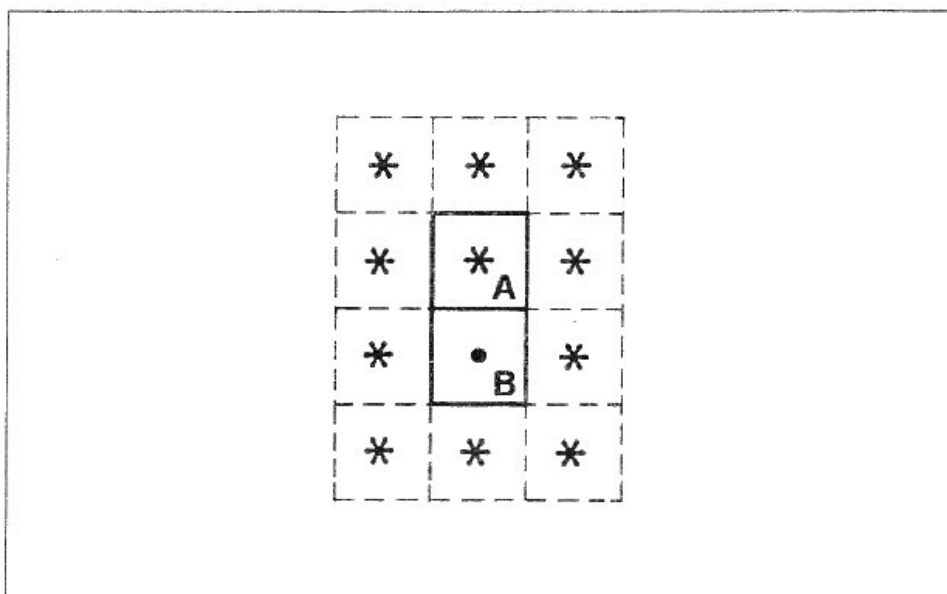
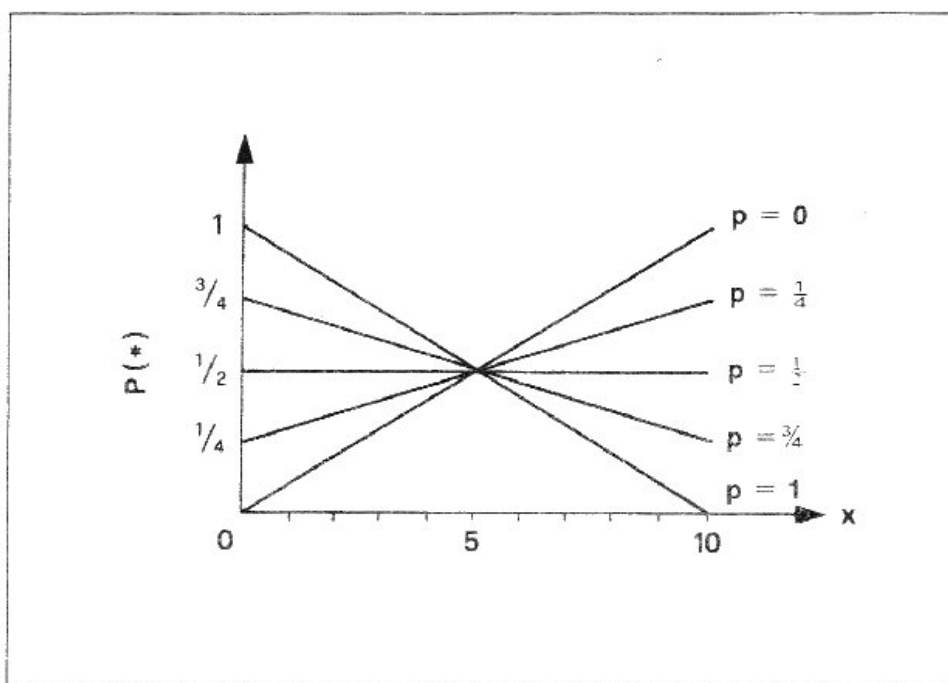



Fig. 19 *Grafico di P[*] = (1-2p)x/10 + p*

la probabilita' di questo evento cresce all'aumentare dei vicini che la pensavano come lui.
Se assegniamo p = 1, si ha P[*] = 1-x/10 e nel caso di fig.18 sara' l'elettore conservatore a trasformarsi in riformista. Per p = 1 il programma OPINIONE si comporta come il programma ANTIVOTERS di cui si parla nella rivista citata.

Riassumendo

INPUT (220-280)
1 dimensioni del rettangolo (trasformato in toro)
2 valore di "p" (P)
3 periodo di stampa W (ogni quante unita' di tempo viene mostra-

ta la situazione).
DETERMINAZIONE SCHEMA INIZIALE (310-360)

CICLO PRINCIPALE

4 scelta di una coppia di elettori contigui di diversa opinione; aggiornamento variabile "tempo" (380-440)
5 calcolo di "x" (450-580) e P[*]
6 determinazione di un numero casuale R=RND(U) ∊ [0,1]: se R>P[*] il riformista si trasforma in conservatore; se R< =P[*] avviene il contrario (590).
7 se tutte le celle contengono lo stesso simbolo si conclude l'elaborazione (380)
8 stampa situazione, se necessario (380)
9 rientro al punto 4.

Per valori p< =1/2, si va verso il "totalitarismo" e, prima o poi, il programma ha termine; per p>1/2, il meccanismo fa si' che gli elettori si distribuiscano equamente tra i due partiti (in media) e l'elaborazione prosegue senza concludersi (tendenzialmente).

Naturalmente le simulazioni piu' "realistiche" si ottengono quando "p" e' 0 o prossimo allo zero.

E' interessante confrontare gli sviluppi nei casi p=1/2 e p=0 (vedi fig.20).

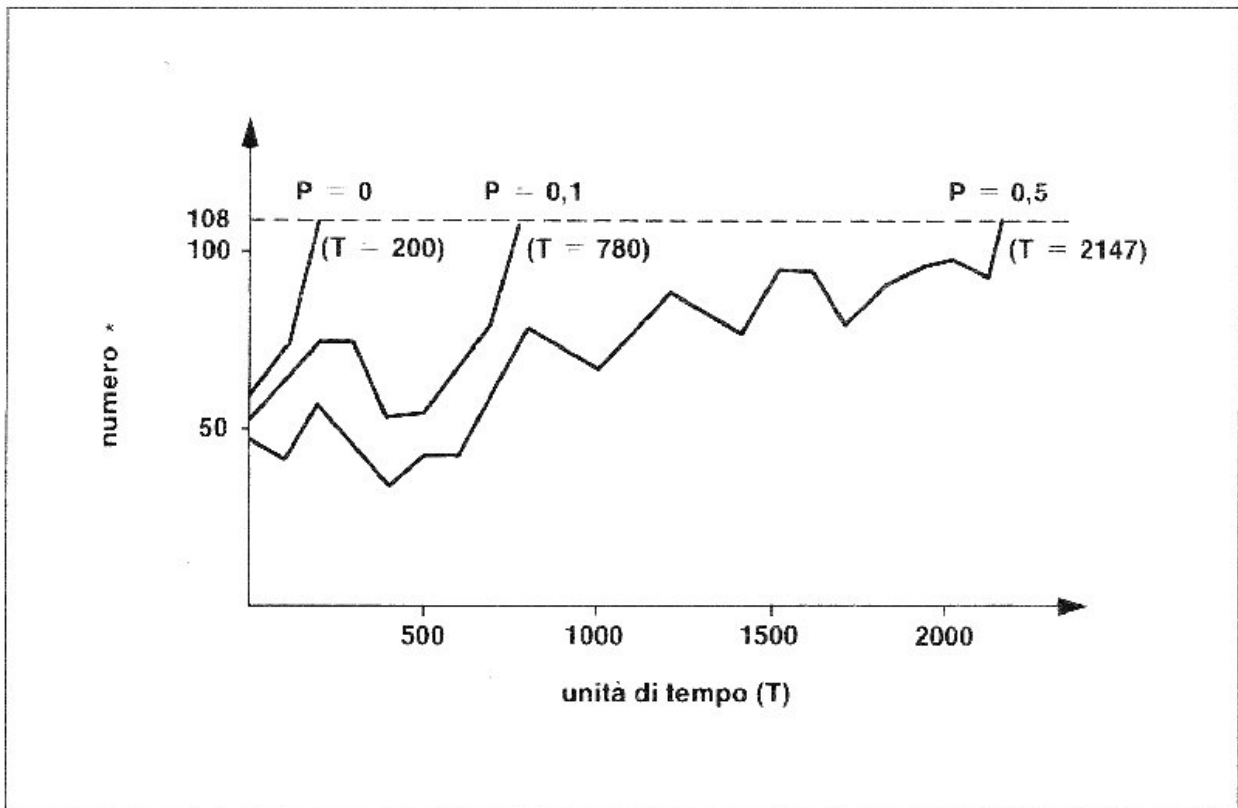



Fig. 20 *L'uniformità viene raggiunta più rapidamente per bassi valori di p*

Quando p = 1/2, i vicini ''non hanno peso'' e si assiste ad una lunga ed estenuante battaglia tra le due fazioni con capovolgimenti di fronte abbastanza frequenti. Lo schema si mantiene irregolare per molto tempo e la convergenza e' piuttosto lenta.

Per p = 0, l'elettore ''stabile'' ha bisogno dell'approvazione degli altri: individui isolati cambiano presto bandiera e i sostenitori dei due partiti si radunano rapidamente nel tentativo di minimizzare la linea di contatto con gli avversari.

Il programma OPINIONE e' presente in due versioni a seconda che si desideri l'output su video (OPINIONE/V) o su stampante (OPINIONE/S).

Soprattutto nel secondo caso, conviene scegliere con attenzione le dimensioni del rettangolo e il periodo di stampa: con rettangoli ''grandi'', l'uniformita' viene raggiunta dopo molto tempo ed e' meglio evitare stampe troppo frequenti.

Un consiglio: utilizzate OPINIONE/S solo dopo numerosi esperimenti con la versione video variando i parametri in ingresso (naturalmente per p > 1/2 bisogna usare OPINIONE/V altrimenti si ottengono stampe a non finire).

Il dischetto contiene, sganciata dai vari menú, la versione compilata del programma 'opinione': OPINIONV.EXE (MS DOS); c/opinione/v (C 64).

ULTERIORI COMMENTI AL LISTATO

Sub. 30-50 e 60-80: calcolo della posizione delle celle modulo L-1 e C-1 (trasformazione del rettangolo in toro)
Sub. 100-210: stampa situazione
220-280: richiesta dati
310-360: schema iniziale
380-440: scelta coppia elettori
450-500: calcolo di ''x''
590-600: risoluzione del confronto tra gli elettori scelti.

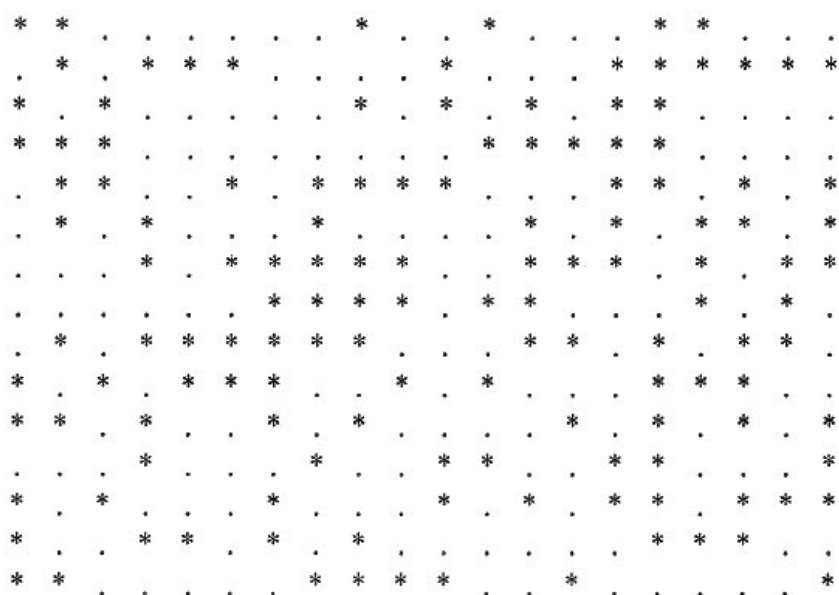
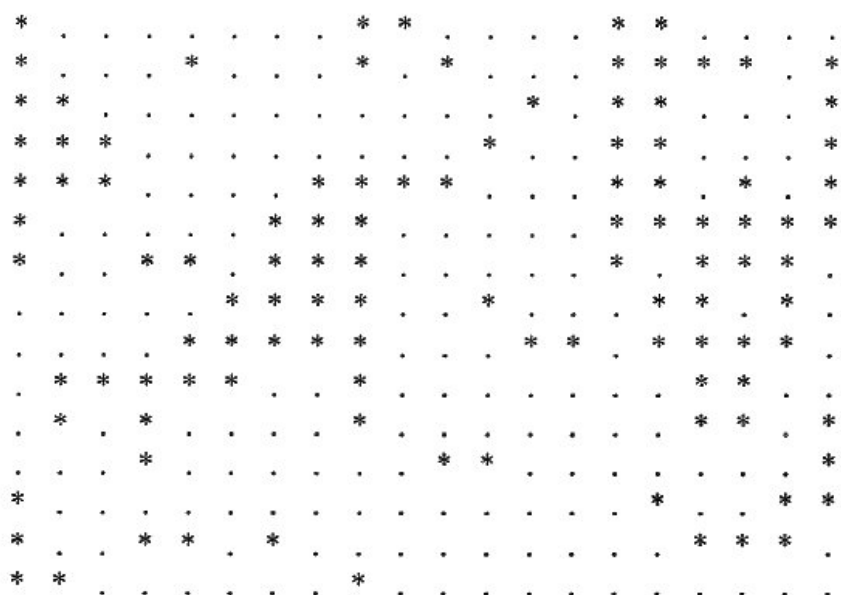
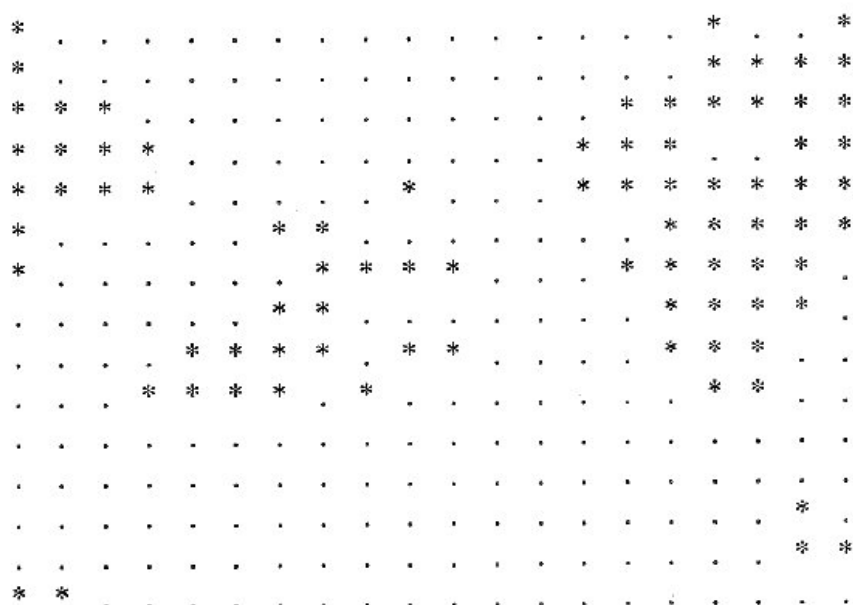
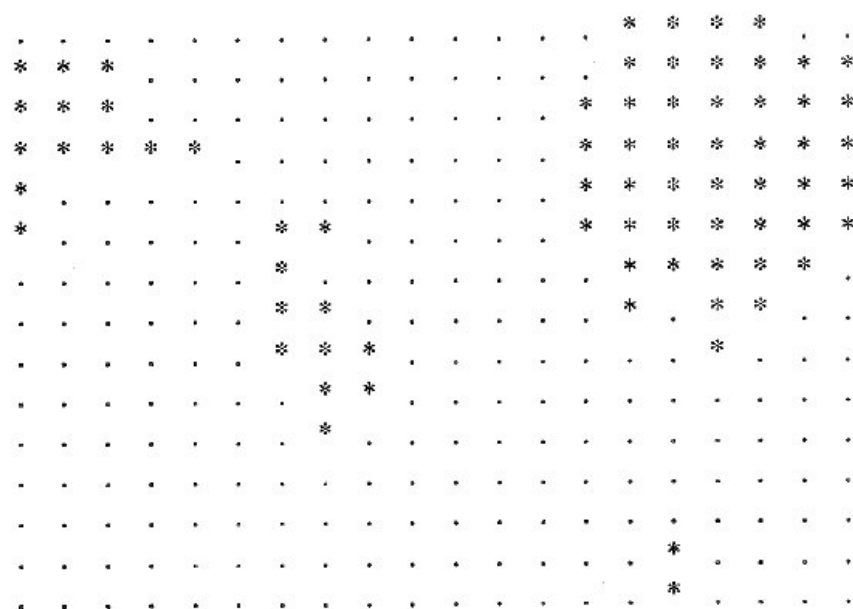
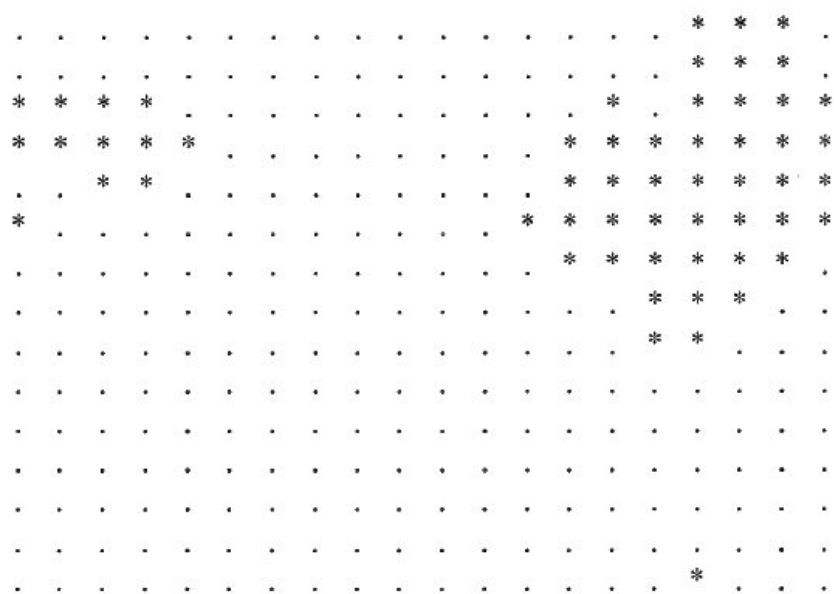
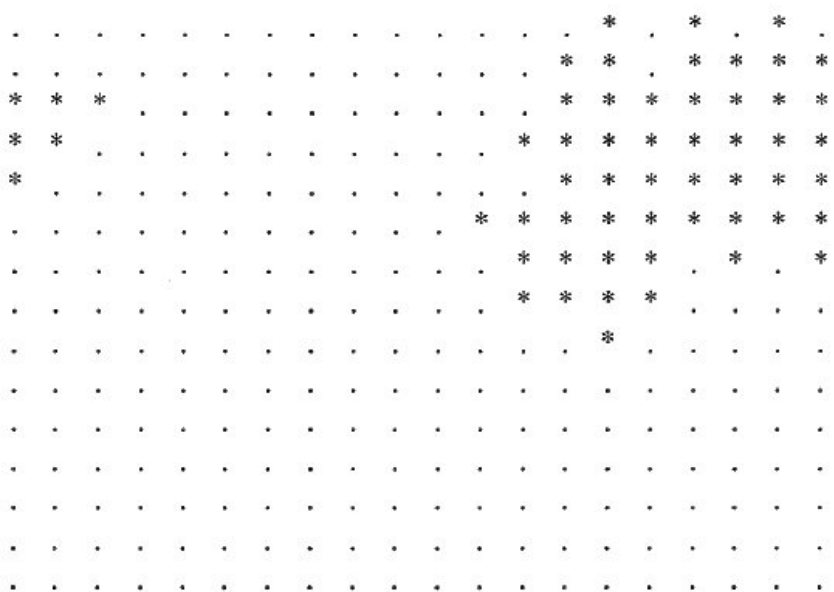
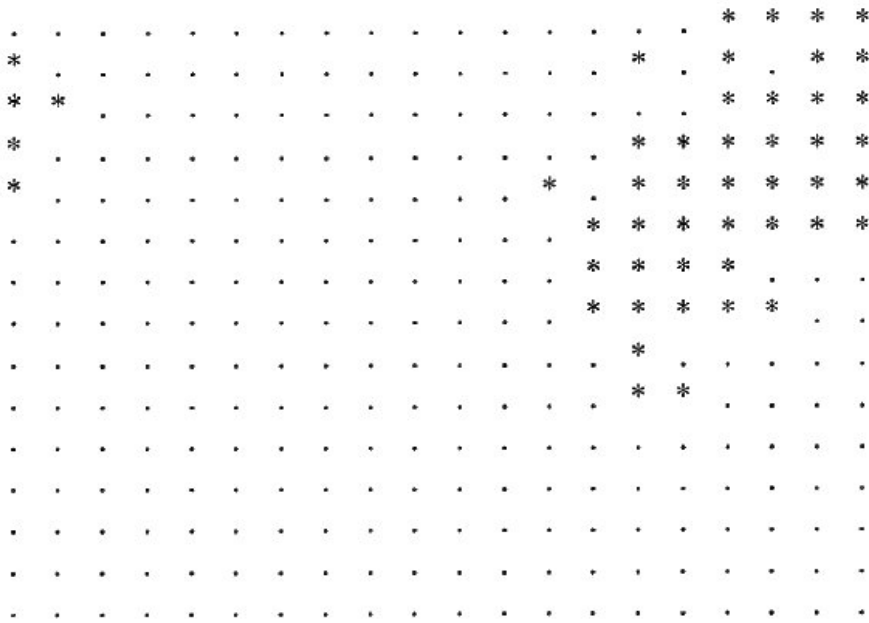
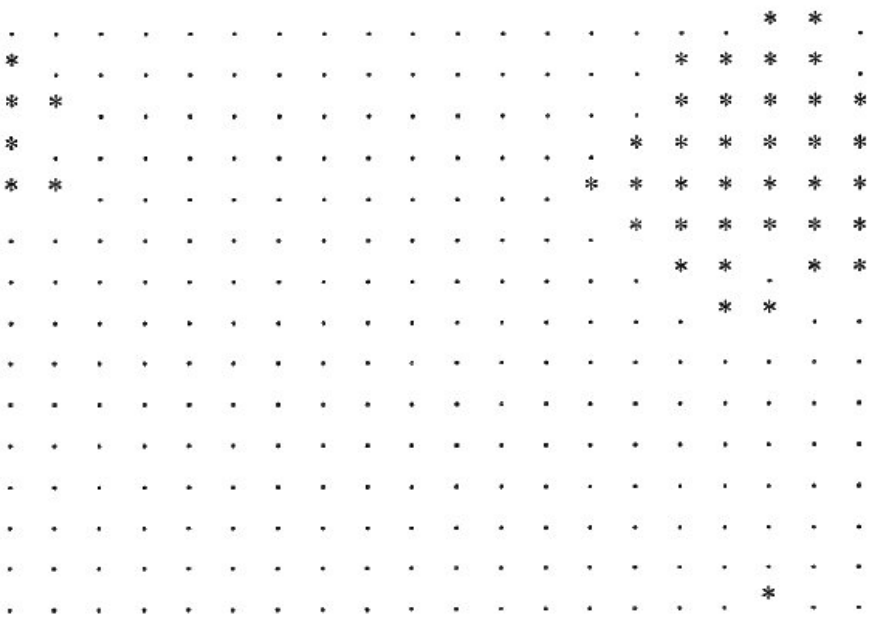
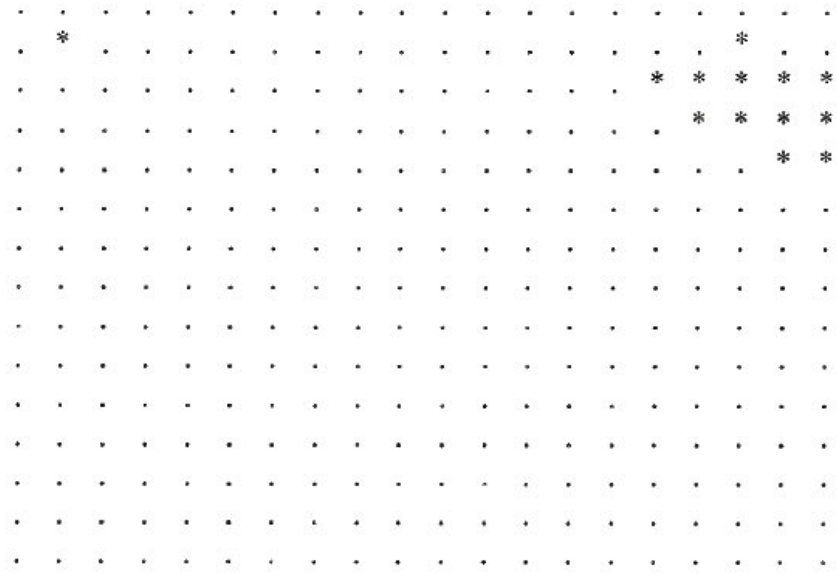
Listato del programma OPINIONE

```
 10 GO TO 220
 30 IF ML>L1 THEN ML=ML-L:RETURN
 40 IF ML<Z THEN ML=ML+L:RETURN
 50 RETURN
 60 IF MC>C1 THEN MC=MC-C:RETURN
 70 IF MC<Z THEN MC=MC+C:RETURN
 80 RETURN
100 PRINT''UNITA' DI TEMPO'';T
```

```
120 FOR I=Z TO L1:FOR J=Z TO C1
140 IF A(I,J)=U THEN A$=A$+P$:GO TO 160
150 A$=A$+V$
160 NEXT J
170 PRINT A$:A$='''':NEXT I
210 PRINT'' *='';NA;'' .='';TC-NA:RETURN
220 INPUT''LINEE,COLONNE'';L,C:L1=L-1:C1=
    C-1:DIM A(L1,C1):TC=L*C
280 INPUT P,W:K=(1-2*P)/10
290 T=0:U=1:Z=0:D=2:MU=-1:P$=''* '':V$=''. ''
310 FOR I=Z TO L1:FOR J=Z TO C1
350 IF RND(1)<.5 THEN A(I,J)=U:NA=NA+U
360 NEXT J,I
370 IF NA=Z OR NA=TC THEN GOSUB 100:END
380 IF T-INT(T/W)*W=Z THEN GOSUB 100
390 T=T+U
400 X=INT(L*RND(U)):Y=INT(C*RND(1))
410 IF RND(U)>.5 THEN 510
420 MC=Y+U:GOSUB 60:IF A(X,Y)=A(X,MC) THEN 400
440 XV=X:YV=MC:S=Z:ML=X-U:GOSUB 30
450 FOR I=MU TO D:MC=Y+I:GOSUB 60:S=S+A(ML,MC):NEXT I
460 ML=X+U:GOSUB 30
470 FOR I=MU TO D:MC=Y+I:GOSUB 60:S=S+A(ML,MC):NEXT I
480 MC=Y-U:GOSUB 60:S=S+A(X,MC)
490 MC=Y+D:GOSUB 60:S=S+A(X,MC)
500 GO TO 590
510 ML=X+U:GOSUB 30:IF A(X,Y)=A(ML,Y) THEN 400
530 XV=ML:YV=Y:S=Z:MC=Y-U:GOSUB 60
540 FOR I=MU TO D:ML=X+I:GOSUB 30:S=S+A(ML,MC):NEXT I
550 MC=Y+U:GOSUB 60
560 FOR I=MU TO D:ML=X+I:GOSUB 30:S=S+A(ML,MC):NEXT I
570 ML=X-U:GOSUB 30:S=S+A(ML,Y)
580 ML=X+D:GOSUB 30:S=S+A(ML,Y)
590 IF RND(U)<(K*S+P) THEN A(X,Y)=U:A(XV,YV)=U:NA=
    NA+U:GO TO 370
600 A(X,Y)=Z:A(XV,YV)=Z:NA=NA-U:GO TO 370
```

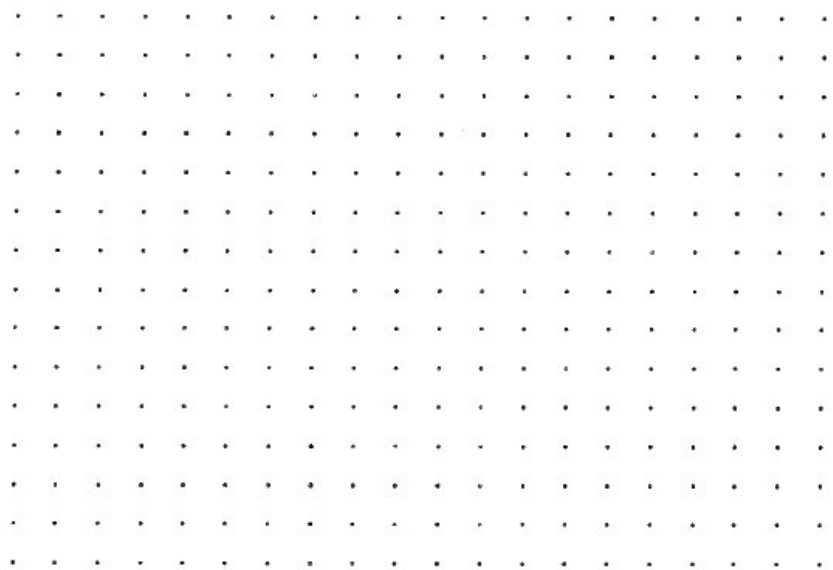
Esempio di output del programma OPINIONE/S

Parametri: L = 15, C = 20, P = 0, W = 100

L'uniformita' viene raggiunta in 831 unita' di tempo.

UNITA' DI TEMPO 0
* = 135 . 165

UNITA' DI TEMPO 100
* = 109 . 191

UNITA' DI TEMPO 200
* = 83 . 217

UNITA' DI TEMPO 300
* = 73 . 227

UNITA' DI TEMPO 400
* = 57 . 243

UNITA' DI TEMPO 500
* = 57 . 243

UNITA' DI TEMPO 600
* = 49 . 251

UNITA' DI TEMPO 700
* = 43 . 257

UNITA' DI TEMPO 800
*= 13 . 287

UNITA' DI TEMPO 831
*= 0 . 300

CAPITOLO 3

ELEMENTI DI STATISTICA

Nel capitolo precedente, problemi di calcolo delle probabilita' di un certo interesse sono stati affrontati con l'ausilio di strumenti matematici relativamente semplici. Un'analoga operazione in relazione alla statistica risulta piu' difficile in quanto problemi di statistica si collocano a diversi e precisi livelli di difficolta' e i piu' interessanti (campionamenti, correlazioni ecc.) richiedono un bagaglio matematico avanzato.

In queste pagine ci si limitera' pertanto agli argomenti di base; il lettore interessato ad approfondire questa parte trovera' in ''Probabilita' Statistica e Termodinamica'' [3] il logico proseguimento e numerosi programmi.

3.1 DATI

Siete stati nominati responsabili della produzione del calzaturificio X che iniziera' l'attivita' nei prossimi mesi. Fra i moltissimi problemi da risolvere dovete naturalmente decidere, per ogni modello di calzatura, quanti paia di scarpe delle varie misure intendete produrre.

Due sono le situazioni che dovete assolutamente evitare: che troppi potenziali clienti non trovino le scarpe adatte ai loro piedi e che nei vostri magazzini si accumulino scarpe che nessuno vuole perche' troppo grandi o troppo piccole.

Per organizzare un piano di produzione sensato (grossolani errori vi costerebbero il posto), vi servono delle informazioni e, se queste non sono gia' disponibili, dovete procurarvele, sintetizzarle ed interpretarle.

Universo statistico

Limitiamoci ad un tipo di calzatura destinata al pubblico maschile e supponiamo che siate interessati al solo mercato nazionale: la popolazione maschile residente in Italia costituira' il vostro UNIVERSO STATISTICO (o COLLETTIVO STATISTICO); le singole persone saranno le UNITA' STATISTICHE.

Carattere

Cio' che intendete esaminare in relazione alle varie unita' statistiche, in questo caso il ''numero di scarpa'', e' il CARATTERE che, in generale, puo' essere espresso da un numero (risultato di un qualche tipo di misura) o dalla presenza o meno di certi attributi. Nel primo caso si parla di caratteri QUANTITATIVI (eta', altezza, peso ecc.), nel secondo di caratteri QUALITATIVI (colore degli occhi, appartenenza ad un certo partito politico ecc.).

Rilevazioni

Fissato l'universo statistico e il carattere, o i caratteri, oggetto di studio, e' necessario procurarsi le informazioni ovvero RILEVARE i DATI. La rilevazione puo' essere TOTALE, se riferita all'intero universo; o PARZIALE, se riferita ad un CAMPIONE ovvero ad un sottoinsieme opportunamente scelto.

Nella vostra situazione dovrete per forza accontentarvi di una rilevazione parziale. Il campione dovra' essere RAPPRESENTATIVO dell'intero universo e di dimensioni tali da consentire l'estensione dei risultati a tutta la popolazione entro ragionevoli margini di sicurezza.

Poiche' una rilevazione su vasta scala puo' essere assai costosa mentre una rilevazione troppo limitata puo' essere assai rischiosa, la determinazione di un campione adatto e' un problema delicato; vi si accennera' nel quarto paragrafo di questo capitolo (Inferenza Statistica).

Supponiamo risolto il problema del campionamento: ora disponete di una consistente mole di dati; pagine fitte di numeri, risultati della vostra indagine statistica.

Sintetizzare i dati

L'informazione che vi serve e' contenuta nei dati in forma, per cosi' dire, polverizzata: per ottenere qualcosa di utile e' necessario un lavoro di sintesi. Si tratta dunque di elaborare i dati in modo da ottenere un insieme limitato di valori che riflettano caratteristiche ritenute importanti.

Abbandoniamo l'esempio del calzaturificio e consideriamo una situazione piu' semplice:

Nella TABELLA 3.1 sono riportati i voti di un compito di matematica svolto nella classe 1' sez. A. Il professore ha raccolto questi dati per cercare una risposta ai seguenti quesiti:
- il testo proposto era adeguato o eccessivamente difficile?
- gli argomenti oggetto della prova sono stati recepiti allo stesso modo da tutti gli studenti?

Tabella 3.1

6.8	8.0	7.1	8.0	5.8	5.5	4.5	5.6	5.1
6.6	6.5	5.5	8.0	5.0	3.8	7.5	5.9	8.0
5.0	7.5	7.0	6.5	6.8	4.5	6.1	6.6	3.9

(Per i voti intermedi si e' posto: 6+ = 6.1; 6 1/2 = 6.5; 6/7 = 6.6 ; 7-- = 6.8; 7- = 6.9 ecc.)

L'universo statitistico e' la classe in esame e le unita' statistiche sono i singoli studenti: la rilevazione e' totale e il carattere esaminato e' il voto ottenuto, carattere quantitativo.

Il numero di dati e' talmente esiguo che, in realta', si potrebbe tentare una risposta semplicemente analizzando la tabella, ma e' facile immaginare situazioni in cui si considerano migliaia di dati la cui elencazione non dice granche'. Inoltre potrebbe risultare utile il confronto con un altro insieme di dati (ad esempio i voti ottenuti in una classe parallela in cui si e' dedicato piu' tempo agli esercizi) e la comparazione dei dati grezzi diverrebbe difficoltosa.

Iniziamo il lavoro di sintesi che, grazie alle modeste dimensione della Tabella 3.1, potremo eseguire dettagliatamente senza ricorrere al calcolatore.

Una semplice riorganizzazione dei dati puo' gia' dire qualcosa; nella Tabella 3.2 i dati sono disposti, per righe, in ordine crescente:

Tabella 3.2

3.8	3.9	4.5	4.5	5.0	5.0	5.1	5.5	5.5
5.6	5.8	5.9	6.1	6.5	6.5	6.6	6.6	6.8
6.8	7.0	7.1	7.5	7.5	8.0	8.0	8.0	8.0

Si nota ora che il voto piu' basso e' 3.8, quello piu' alto e' 8. Se raggruppiamo i dati secondo la seguente

Tabella 3.3

i voti x :	sono
$3 \leqslant x < 4$	2
$4 \leqslant x < 5$	2
$5 \leqslant x < 6$	8
$6 \leqslant x < 7$	7
$7 \leqslant x < 8$	4
$8 \leqslant x < 9$	4

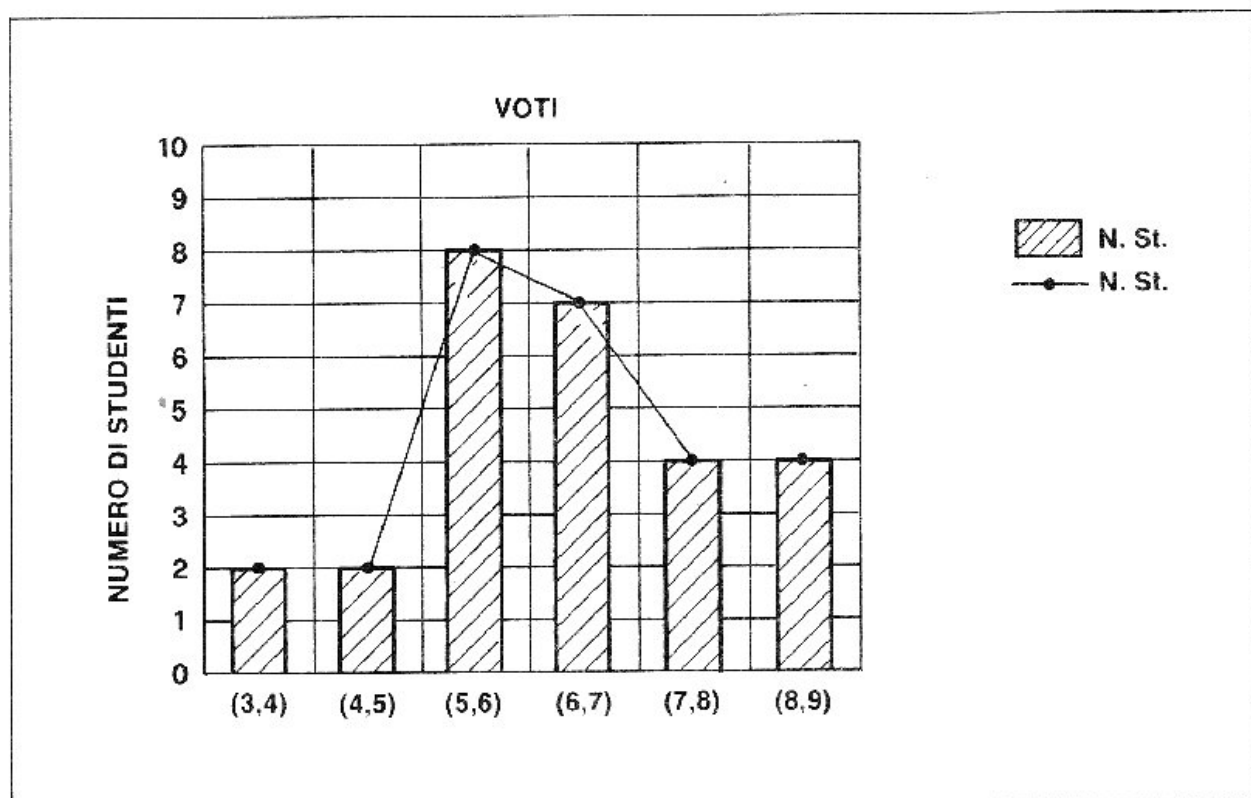



Fig. 21 *Rappresentazione grafica dei dati della Tabella 3.1*

e riportiamo i valori su un grafico (fig.21), saltano all'occhio altre caratteristiche: i voti estremi sono piu' rari mentre la maggioranza si concentra tra il 5 e il 7.

Vi sono molti modi per rappresentare un insieme di dati ed esistono in commercio programmi che forniscono rappresentazioni grafiche veramente suggestive (vedi fig. 22 e 23).

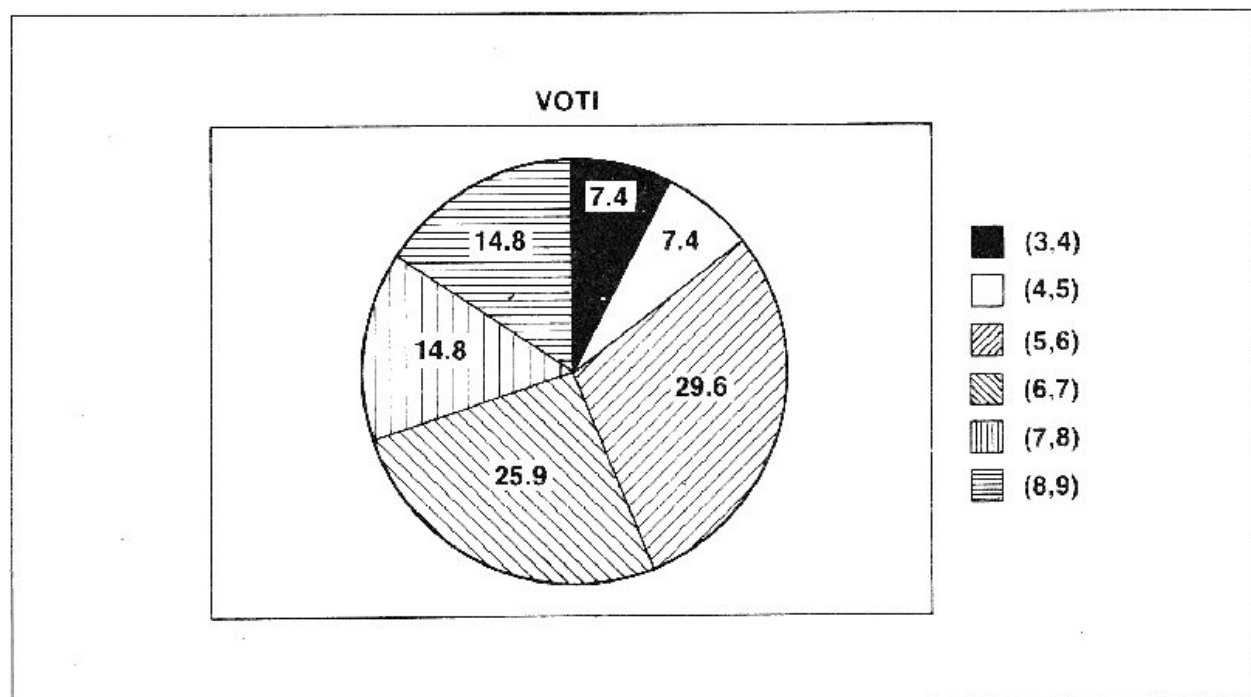



Fig. 22 *Rappresentazione grafica dei dati della Tabella 3.1*

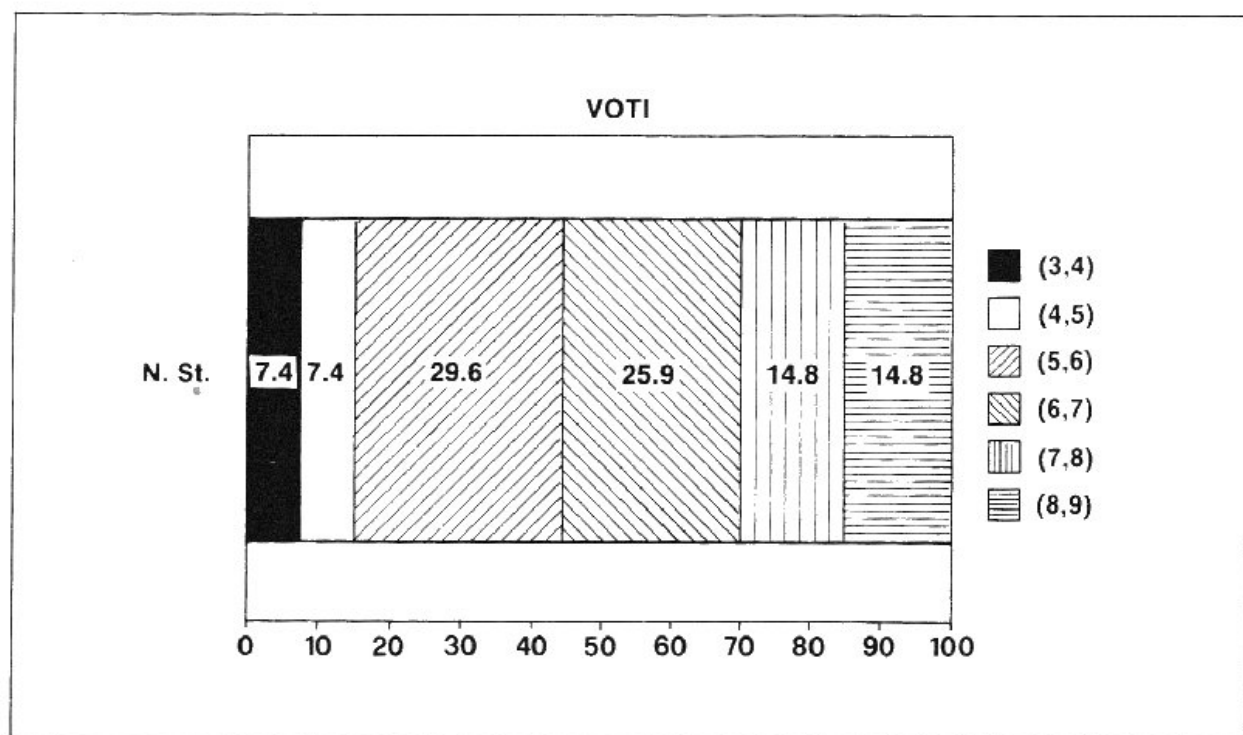



Fig. 23 *Rappresentazione ~~a torta~~ dei dati della Tabella 3.1*

3.2 INDICI SINTETICI DI VARIABILI SEMPLICI: INDICI DI POSIZIONE

Consideriamo un insieme di dati relativi ad un carattere quantitativo; indichiamo con "n" il loro numero e con x_i $i=1,2,\ldots,n$ i valori espressi, in generale, da numeri reali (nell'esempio precedente e' $n = 27$ = numero dei voti e x_i = voto).

Al di la' delle possibili rappresentazioni grafiche, e' molto utile calcolare degli INDICI numerici che evidenzino certe caratteristiche dell'intero insieme; se un grafico fornisce una visione globale, un numero si presta ad essere confrontato con altri numeri e, come si vedra', la cosa puo' essere molto utile.

Gli indici statistici piu' frequentemente utilizzati si distinguono in indici di POSIZIONE e indici di DISPERSIONE; i primi forniscono informazioni sulla "grandezza" dei dati, i secondi sulla loro diversita' reciproca.

Media aritmetica

L'indice di posizione piu' famoso e' senz'altro la MEDIA ARITMETICA: "la media aritmetica di "n" dati e' quel valore che, sostituito a ciascuno dei dati, ne lascia invariata la somma".

Detto M tale valore, e', per definizione,

$$x_1 + x_2 + \ldots + x_n = M + M + \ldots + M = nM$$

da cui

$$M = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_i x_i$$

ESERCIZIO

Mostrare che, se x' e x'' sono rispettivamente il minimo e il massimo di x_i, e' x' $\leqslant$ M $\leqslant$ x''.

Una proprieta' caratteristica della media aritmetica e':

$$\sum_i (x_i - M) = 0$$

ovvero ''la somma degli scarti dalla media aritmetica e' zero''.

In effetti:
$(x_1 - M) + (x_2 - M) + \ldots + (x_n - M) =$
$= (x_1 + x_2 + \ldots + x_n) - (M + M + \ldots + M) = nM - nM = 0.$

Per quanto riguarda la Tab. 3.1 si ha M = 167.1/27 = 6.189

Media aritmetica ponderata

Quando i dati sono raggruppati e il dato x_i compare nella statistica f_i volte, l'espressione di M e' data da:

$$M = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_k x_k}{f_1 + f_2 + \ldots + f_k}$$

dove ''k'' e' il numero di raggruppamenti (media aritmetica PONDERATA).

Volendo applicare la formula precedente ai dati della Tab. 3.3, potremo scegliere come x_i il valore centrale di ogni intervallo della suddivisione scelta (cioe' considerare uguali a 8.5 tutti i valori compresi tra 8 e 9 ecc.), ottenendo:

$$M = \frac{3.5x2 + 4.5x2 + 5.5x8 + 6.5x7 + 7.5x4 + 8.5x4}{2+2+8+7+4+4} =$$

$$= 169.5/27 = 6.278$$

La media ottenuta in questo modo e' diversa dalla precedente poiche', considerando il valore centrale, si commette un errore di approssimazione.

Alla suddivisione in classi e alla media ponderata si ricorre soprattutto quando il carattere quantitativo e' CONTINUO ovvero quando il dato puo' assumere, almeno teoricamente, un qualsiasi valore reale di un certo intervallo. Cosi', se viene misurata la statura di un gruppo di persone, converra' contare il numero di indi-

vidui la cui altezza e' compresa tra 175 e 180 cm, tra 180 e 185 cm ecc. evitando, gia' in fase di raccolta dei dati, una eccessiva frantumazione. Diverso il discorso per un carattere DISCRETO come, ad esempio, il numero di persone che costituiscono un nucleo familiare; una tabella del tipo:

i nuclei familiari formati da	sono
1 persona	10
2 persone	52
3 persone	60
4 persone	47
ecc.	

contiene dati raggruppati per classi, ma, poiche' si tratta di un carattere discreto, la media aritmetica coincide con quella ponderata.

Nel caso del voto, il carattere andrebbe considerato come discreto, dato che esiste un numero finito di voti intermedi convenzionalmente assegnati (7+, 7 1/2 ecc.), ma la suddivisione in classi di ampiezza "1 voto" risulta comoda sia dal punto di vista grafico sia per gli scopi che si propone il professore.

Mediana

Un altro indice di posizione di uso molto frequente e' la MEDIANA (λ) definita come "valore centrale" dei dati quando questi siano ordinati in ordine non decrescente.

Se $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n$ e'

$\lambda = x_{(n+1)/2}$ se n e' dispari (valore centrale)

mentre si assume

$\lambda = \frac{1}{2} [x_{n/2} + x_{(n/2)+1}]$ se n e' pari (media aritmetica dei 2 valori centrali)

ESEMPIO 1 :

DATI : 7; 2; 7; 10; 100 (n = 5)

DATI ORDINATI : 2; 7; 7; 10; 100

⇧

$(n+1)/2 = 3; \lambda = 7$

ESEMPIO 2 :

DATI : 4; 1; 1; 6; 30; 5 (n = 6)

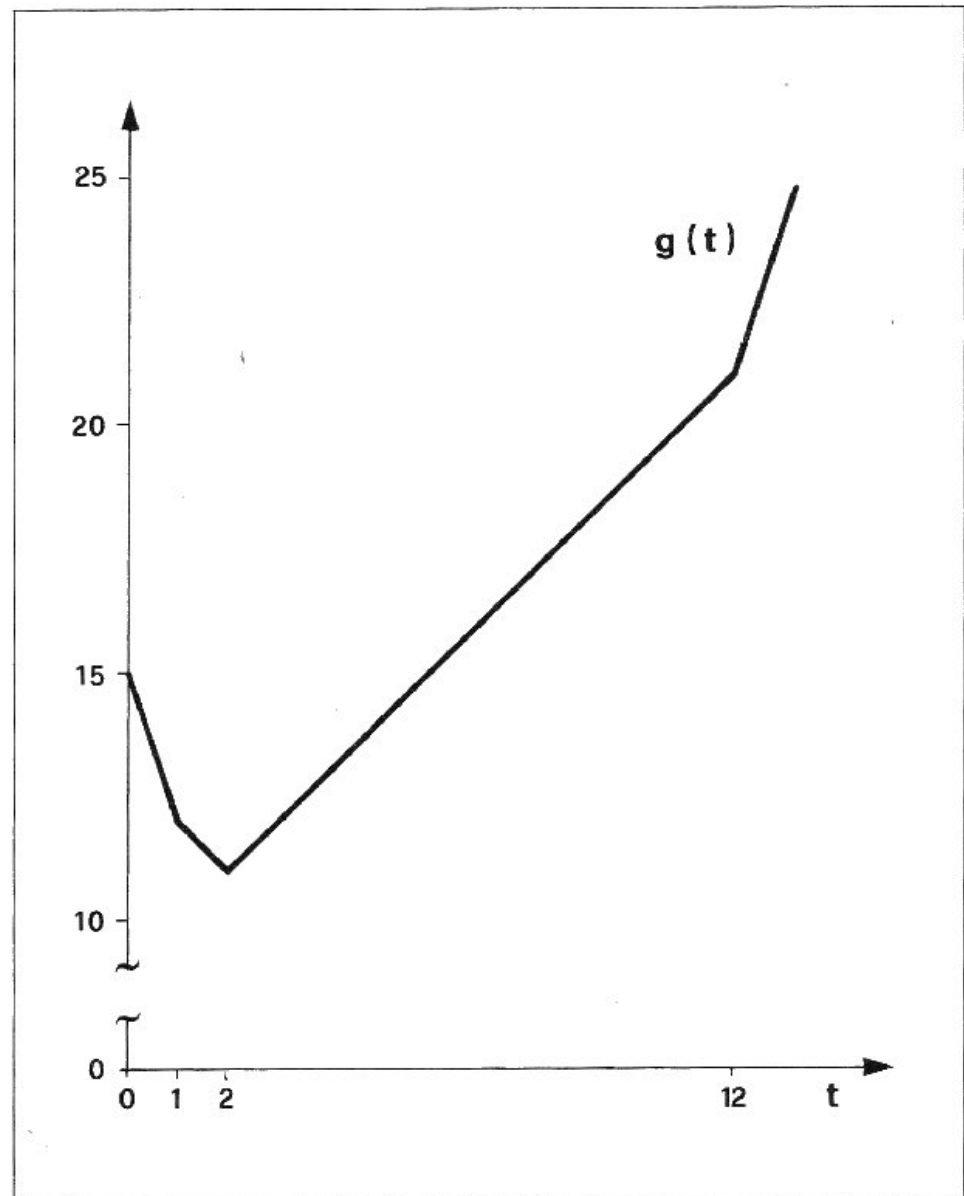
DATI ORDINATI : 1; 1; 4; 5; 6; 30

⇧ ⇧

$\lambda = (4 + 5)/2 = 4.5$

Nell'esempio dei voti e' n = 27 e la mediana sara' il 14' valore delle Tab. 3.2 : $\lambda = 6.5$ il che' vuol dire che una meta' della classe ha ottenuto voti inferiori (o uguali) a 6.5, l'altra meta' superiori (o uguali) a 6.5.

Una grande differenza tra media e mediana sta ad indicare una distribuzione ''sbilanciata'': un solo valore eccezionalmente grande o eccezionalmente piccolo influenza la media aritmetica ma non la mediana (Es. DATI 1; 2; 96 : M = 33; $\lambda = 2$).

Fig. 24 *Grafico di* $g(t) = |1-t| + |2-t| + |12-t|$

Media aritmetica e mediana godono di notevoli proprieta': la media aritmetica rende minima la funzione $f(t) = \sum_i (x_i-t)^2$; la mediana minimizza la funzione $g(t) = \sum_i |x_i-t|$. Verifichiamole in un esempio:

Sia $n = 3 : x_1 = 1; x_2 = 2; x_3 = 12$. La media aritmetica e' 5, la mediana e' 2.

La funzione

$$f(t) = (1-t)^2 + (2-t)^2 + (12-t)^2 = 3t^2-30t + 149$$

assume il valore minimo per $t = 5$, ascissa del vertice della parabola.

La funzione

$$g(t) = |1-t| + |2-t| + |12-t|$$

puo' essere riscritta come

$$g(t) = \begin{cases} 15-3t & \text{per } t<1 \\ 13-t & \text{per } 1 \leqslant t<2 \\ t+9 & \text{per } 2 \leqslant t<12 \\ 3t-15 & \text{per } t \geqslant 12 \end{cases}$$

il minimo si ottiene (vedi fig.24) per $t = 2$.

Quartili

In analogia con la mediana, che divide i dati ordinati in due gruppi ugualmente numerosi, si possono definire altri indici di posizione. Per primo, secondo, terzo QUARTILE (q_1, q_2, q_3) si intendono i tre valori che dividono i dati in quattro tronconi di pari dimensione, una volta che i dati siano disposti in ordine non decrescente.

In riferimento alla Tab. 3.2, il 1', 2', 3' quartile corrispondono rispettivamente al 7', 14', 21' valore dell'elenco:

$$q_1 = 5.1 \; ; \; q_2 = 6.5 \; ; \; q_3 = 7.1$$

il secondo quartile (q_2) e' la mediana (λ).

Generalizzando ulteriormente, si possono definire analoghi indici di posizione (quintili, decili, percentili) di un certo interesse se i dati sono molto numerosi.

Moda

Definiamo come MODA (o VALORE NORMALE) il dato che compare piu' spesso nell'insieme dei dati. Se il carattere in esame e' di tipo qualitativo o quantitativo discreto, la moda sara' appunto la caratteristica o il valore cui corrisponde la massima frequenza; nel caso di un carattere continuo (o, comunque, registrato per raggruppamenti), avra' senso parlare di CLASSE MODALE in riferimento all'intervallo cui compete la maggior frequenza (suppo-

nendo di considerare intervalli della stessa ampiezza). Nell'esempio dei voti, la moda e' 8.

Oltre la media aritmetica, sono definite altre medie utilizzate per specifici problemi:

Media geometrica

si definisce MEDIA GEOMETRICA dei numeri positivi $x_1, x_2, .. ,x_n$ quel numero G che, sostituito ai dati, ne lascia invariato il prodotto.

Per definizione:

$$x_1 \cdot x_2 \ldots \cdot x_n = G \cdot G \ldots \cdot G = G^n$$

da cui

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \quad .$$

Il calcolo di G viene normalmente eseguito utilizzando i logaritmi:

$$\log G = \frac{1}{n} [\log x_1 + \log x_2 + \ldots + \log x_n]$$

(il logaritmo delle media geometrica e' uguale alla media aritmetica dei logaritmi dei dati).

Sempre in relazione ad un insieme di dati positivi $x_1, x_2, .. ,x_n$, ricordiamo la definizione di altre due medie (raramente utilizzate):

- MEDIA QUADRATICA :

$$Q = \sqrt{\frac{x^2_1 + \ldots + x^2_n}{n}}$$

- MEDIA ARMONICA :

$$A = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

E' possibile dimostrare che, se $x_i > 0 \ \forall \ i$, e'
$A \leqslant G \leqslant M \leqslant Q$.

Esercizio. Dati i valori 3; 3; 5; 7; 10, verificare che $A \cong 4.51$; $G \cong 5.01$; $M = 5.6$; $Q \cong 6.2$.

Esercizio. Verificare che, se gli n dati sono tutti uguali ad uno stesso numero $h > 0$, e' $A = G = M = Q = h$.

Esercizio. Dimostrare che $A \leqslant G \leqslant M \leqslant Q$ nel caso di due dati positivi.

Soluzione

1\. E' $(a - b)^2 ab \geqslant 0$

$$[(a + b)^2 - 4ab]\, ab \geqslant 0$$

$$(a + b)^2 ab \geqslant 4a^2 b^2$$

$$(a + b) \sqrt{ab} \geqslant 2ab$$

$$\sqrt{ab} \geqslant \frac{2ab}{a + b} = \frac{2}{\frac{1}{a} + \frac{1}{b}}$$

cioe' $G(a,b) \geqslant A(a,b)$

2\. E' $(\sqrt{a} - \sqrt{b})^2 \geqslant 0$

$$a + b - 2\sqrt{ab} \geqslant 0$$

$$a + b \geqslant 2\sqrt{ab}$$

$$\frac{a + b}{2} \geqslant \sqrt{ab}$$

quindi $M(a,b) \geqslant G(a,b)$

3\. E' $\frac{(a - b)^2}{4} \geqslant 0$

$$\frac{a^2 + b^2 - 2ab}{4} \geqslant 0$$

$$\frac{2a^2 + 2b^2 - a^2 - b^2 - 2ab}{4} \geqslant 0$$

$$\frac{a^2 + b^2}{2} - \frac{a^2 + b^2 + 2ab}{4} \geqslant 0$$

$$\frac{a^2 + b^2}{2} \geqslant \left(\frac{a + b}{2}\right)^2$$

$$\sqrt{\frac{a^2 + b^2}{2}} \geqslant \frac{a + b}{2}$$

ovvero $Q(a,b) \geqslant M(a,b)$

Chi si avvicina per la prima volta alla statistica sara', a questo punto, un po' disorientato: si voleva sintetizzare un insieme di dati e ci si trova di fronte ad una miriade di indici. Il fatto e' che ciascuno degli indici introdotti evidenzia certe caratteristiche dei dati; alcune informazioni potranno risultare piu' utili, altre meno. Se, tornando all'esempio del calzaturificio, decidete di produrre inizialmente uno stock di scarpe di un'unica misura (tanto per saggiare il gradimento del pubblico), vi interessera' la moda piu' della media o della mediana.

Un esempio

Concludiamo questo paragrafo con un'applicazione della media geometrica (l'esempio e' preso da [12], pag. 16):

La popolazione di una citta' negli anni dal 1980 al 1984 e'

Anno	popolazione
1980	40000
1981	42000
1982	50400
1983	57960
1984	63756

i tassi di incremento annuali sono:

80-81	42000/40000 = 1.05	(a_1)
81-82	50400/42000 = 1.20	(a_2)
82-83	57960/50400 = 1.15	(a_3)
83-84	63756/57960 = 1.10	(a_4)

Definiamo come ''tasso medio di crescita'' nel quadriennio quel tasso ''t'' che applicato ripetutamente a partire dal 1980 riproduca lo stesso ammontare della popolazione nel 1984. ''t'' e' la media geometrica dei tassi di incremento annuali:

$$t = \sqrt[4]{1.05 \times 1.2 \times 1.15 \times 1.1} = \sqrt[4]{1.5939} = 1.1236091$$

infatti:
40000 x t = 44944
44944 x t = 50500
50500 x t = 56742
56742 x t = 63756

In effetti: $P_{81} = a_1P_{80}$; $P_{82} = a_2P_{81}$; $P_{83} = a_3P_{82}$; $P_{84} = a_4P_{83}$. Sostituendo, $P_{84} = a_4a_3a_2a_1P_{80}$. Per definizione di di ''t'', deve essere $P_{81} = tP_{80}$; $P_{82} = tP_{81}$ ecc. in modo che $P_{84} = t \cdot t \cdot t \cdot t \cdot P_{80}$. Quindi

$t^4 = a_1a_2a_3a_4$ e $t = \sqrt[4]{a_1a_2a_3a_4}$.

3.3 INDICI SINTETICI DI VARIABILI SEMPLICI: INDICI DI DISPERSIONE

I voti dei compiti di matematica di Massimo sono 6, 8, 6, 4; quelli dei compiti di Elisabetta 6, 6, 6, 6. Per entrambi e' M = 6, Mediana = 6, Moda = 6 ma, mentre Elisabetta e' sempre sufficientemente preparata, il rendimento di Massimo e' discontinuo.

Variabilità dei dati

La DISPERSIONE (o VARIABILITA') di un insieme di dati e' senz'altro un aspetto importante e risultera' utile calcolare dei valori che ne misurino l'entita' (INDICI DI DISPERSIONE).

Una prima informazione e' data dal CAMPO DI VARIAZIONE (V) definito come differenza tra il maggiore e il minore dei dati; se

$x' = \min x_i$ e
$x'' = \max x_i$ e'
$V = x'' - x'$.

Generalmente non si tratta di un valore molto indicativo (un solo dato eccezionalmente grande o eccezionalmente piccolo ''dilata'' il campo di variazione); puo' rivelarsi utile se V risulta particolarmente ridotto rispetto all'ordine di grandezza dei dati.

Il campo di variazione puo' essere espresso attraverso la coppia (x',x'') in modo da fornire anche indicazioni ''di posizione'': tutti i dati sono compresi nell'intervallo (x',x'').

Scostamento semplice medio

Piu' interessante e' lo SCOSTAMENTO SEMPLICE MEDIO definito come:

$$\sigma' = \frac{1}{n} \sum_i |x_i - M|$$

che dice di quanto, mediamente, i dati si discostano dalla loro media aritmetica.

In relazione all'ultimo esempio si ha:
- Voti di Massimo ($M = 6$):
  $(x',x'') = (4,8)$; $V = 8 - 4 = 4$

$$\sigma' = \frac{1}{4} [|6 - 6| + |8 - 6| + |6 - 6| + |4 - 6|] = 1$$

- Voti di Elisabetta ($M = 6$):
  $(x',x'') = (6,6)$; $V = 6 - 6 = 0$
  $\sigma' = 0$.

Scostamento quadratico medio

Piu' comodo da calcolare e di uso piu' comune e' lo SCOSTAMENTO QUADRATICO MEDIO (o DEVIAZIONE STANDARD):

$$\sigma = \sqrt{\frac{1}{n} \sum_i (x_i - M)^2}$$

Varianza

il cui significato e' analogo a quello di σ' (anche se, nel computo

di σ, grossi scarti hanno ''maggiore peso''). Il quadrato dello scostamento quadratico medio (σ^2) e' la VARIANZA.

Per quanto riguarda i voti di Massimo abbiamo:

x_i	x_i - M	$(x_i - M)^2$
6	0	0
8	2	4
6	0	0
4	-2	4

la somma dei valori dell'ultima colonna e' 8 per cui

$$\sigma = \sqrt{\frac{1}{4} \cdot 8} = \sqrt{2} = 1.414.$$

Ovviamente σ (Voti di Elisabetta) = 0.

Esercizio. Dimostrare che

$$\sum_i (x_i - M)^2 = \sum x^2_i - nM^2$$

Soluzione:

$$\sum_i (x_i - M)^2 = \sum_i (x^2_i - 2Mx_i + M^2) =$$

$$= \sum_i x^2_i - 2M \sum_i x_i + nM^2 =$$

$$= \sum_i x^2_i - 2M(Mn) + nM^2 = \sum_i x^2_i - nM^2.$$

Questa identita' consente di calcolare σ come

$$\sigma = \sqrt{\frac{1}{n} [\sum_i x^2_i - nM^2]} = \sqrt{\frac{\sum_i x^2_i}{n} - M^2}$$

una formula alternativa assai utile dato che non e' necessario disporre preventivamente della media aritmetica (si evita di passare in rassegna i dati due volte e quindi di doverli memorizzare in un vettore).

Ricalcoliamo σ (Voti di Massimo) con la nuova formula:

x_i : 6 8 6 4

x^2_i : 36 64 36 16

$\sum_i x_i^2 = 152$; $M^2 = 36$; $n = 4$ per cui

$$\sigma = \sqrt{\frac{152}{4} - 36} = \sqrt{2} \quad .$$

x_i	$\vert x_i - M \vert$	$(x_i - M)^2$
3.8	2.389	5.707
3.9	2.289	5.240
4.5	1.689	2.853
4.5	1.689	2.853
5.0	1.189	1.414
5.0	1.189	1.414
5.1	1.089	1.186
5.5	0.689	0.475
5.5	0.689	0.475
5.6	0.589	0.347
5.8	0.389	0.151
5.9	0.289	0.084
6.1	0.089	0.008
6.5	0.311	0.097
6.5	0.311	0.097
6.6	0.411	0.169
6.6	0.411	0.169
6.8	0.611	0.373
6.8	0.611	0.373
7.0	0.811	0.658
7.1	0.911	0.830
7.5	1.311	1.719
7.5	1.311	1.719
8.0	1.811	3.280
8.0	1.811	3.280
8.0	1.811	3.280
8.0	1.811	3.280
Totale	**28.511**	**41.531**

Campo di variazione = 8 - 3.8 = 4.2
Scostamento semplice medio σ' = 28.511/27 = 1.056
Deviazione standard $\sigma = \sqrt{41.531/27} = 1.24$
Differenza interquartile = 7.1 - 5.1 = 2

Ecco un breve programma Basic per il calcolo di M e σ :

```
10 INPUT''QUANTI DATI'';N
20 FOR I = 1 TO N
30 INPUT''DATO..'';X
40 S=S+X:Q=Q+X*X
50 NEXT I
60 M=S/N:D=SQR(Q/N-M*M)
70 PRINT M,D
```

Differenza interquartile

Altro indice di dispersione e' la DIFFERENZA INTERQUARTILE

$q_3 - q_1$

che, come il campo di variazione, puo' essere espressa dalla coppia (q_1,q_3). Il significato e' semplice: meta' dei dati cadono in questo intervallo.

Tornando all'esempio della Tab. 3.2 (n = 27 voti; M = 6.189), calcoliamo gli indici di dispersione:

Ricapitolando:

CLASSE 1' SEZ. A

M	= 6.189	media aritmetica
λ	= 6.5	mediana
MD	= 8	moda
(x',x'')	= (3.8 , 8)	campo di variazione
V	= 4.2	campo di variazione
(q_1,q_3)	= (5.1 , 7.1)	differenza interquartile
$q_3 - q_1$	= 2	differenza interquartile
σ'	= 1.056	scostamento semplice medio
σ	= 1.24	scostamento quadratico medio
σ^2	= 1.5376	varianza

La statistica mostra che il compito proposto puo' ritenersi adeguato (M e λ oltre ''la sufficienza''); la classe e' abbastanza omogenea: anche se V e' piuttosto elevato, σ' e σ mostrano che, mediamente, ci si discosta di 1 voto dalla sufficienza.

Nella classe parallela, la 1' B, gli studenti hanno usufruito di un'ora di lezione in piu' dedicata al ripasso. I risultati, relativi allo stesso compito valutato con gli stessi criteri, sono riportati nella Tabella 3.4

Tabella 3.4

3.8	4.1	4.5	4.9	5.5	5.8	6.0	6.1	6.5
6.6	6.6	6.6	6.8	6.8	6.8	6.8	6.9	7.0
7.0	7.0	7.0	7.0	7.0	7.1	7.1	7.1	7.6

Per questa classe abbiamo:

CLASSE 1' SEZ. B

M	= 6.37	media aritmetica
λ	= 6.8	mediana
MD	= 7	moda
(x',x'')	= (3.8 , 7.6)	campo di variazione
V	= 3.8	campo di variazione
(q_1,q_3)	= (6 , 7)	differenza interquartile
$q_3 - q_1$	= 1	differenza interquartile
σ'	= 0.76	scostamento semplice medio
σ	= 0.97	scostamento quadratico medio
σ^2	= 0.94	varianza

Dal confronto delle due statistiche si nota come la situazione della 1' B sia decisamente migliore: voti complessivamente piu' alti e maggiore omogeneita'. Sembra proprio che la lezione aggiuntiva sia servita.

Naturalmente, trattandosi di valutazioni scolastiche, il ''6'' risulta un valore particolare poiche' discrimina tra prove sufficienti ed insufficienti; nella 1' A sono pienamente sufficienti ($x_i \geqslant 6$) 15 studenti su 27 (circa il 56%), nella 1' B 21 su 27 (circa il 78%).

Il programma STATISTICA

Il programma STATISTICA, che si propone scopi puramente didattici, consente di calcolare gli indici statistici descritti in relazione ad un insieme di dati positivi che possono essere registrati su disco (o nastro) e da li' richiamati.

Esempio di Output del programma STATISTICA

(Dati: 7, 3, 10, 10, 10, 9, 9)

Risultati :

Media Aritmetica	8.28571429
Media Geometrica	7.76426976
Media Quadratica	8.61891608
Media Armonica	7.01112878
Moda	10

Frequenza val.modale	3
Mediana	9
Primo Quartile	7
Terzo Quartile	10
Campo di variazione	7
Minimo Xi	3
Massimo Xi	10
Differenza interquar.	3
Scost. semplice medio	1.87755102
Scarto quadr. medio	2.3733211
Varianza	5.63265305

3.4. INFERENZA STATISTICA

Nei paragrafi precedenti ci siamo occupati della rappresentazione e della sintesi di un insieme di dati (STATISTICA DESCRITTIVA) indipendentemente dal fatto che questi provenissero dall'intera popolazione o da una sua parte.

E' facile rendersi conto che una rilevazione totale e' molto spesso da escludere: non appena l'universo statistico supera una certa dimensione, diviene praticamente impossibile rilevare i dati da tutte le unita' statistiche. Inoltre una rilevazione su vasta scala puo' essere assai costosa sia in termini monetari (si pensi di valutare l'efficienza di un farmaco) sia in termini di tempo (nei sondaggi preelettorali, i risultati servono subito); per non parlare dei rilevamenti "distruttivi" in cui l'unita' esaminata risulta poi inservibile (per saggiare la resistenza dei tondini metallici utilizzati in edilizia, li si sottopone a "prove a trazione fino a rottura").

Quando si e' costretti ad un rilevamento parziale, si pone un problema: fino a che punto i risultati dedotti dal campione (normalmente assai ridotto rispetto all'universo statistico) possono estendersi alla totalita'?

Di questi tipi di problemi si occupa l'INFERENZA STATISTICA, il cui studio richiede conoscenze matematiche piuttosto avanzate; ci limiteremo a qualche considerazione.

Un primo grosso problema consiste nella determinazione del campione da utilizzare: come sceglierlo e quanto grande ?

Campionamento casuale semplice

Per la scelta, si ricorre spesso a metodi probabilistici; nel CAMPIONAMENTO CASUALE SEMPLICE le "k" unita' che formano il campione vengono estratte casualmente dalle "n" che costituiscono l'universo.

Campioni stratificati

In questo modo si evita di privilegiare, magari inconsciamente, certe categorie e di lavorare su insiemi non rappresentativi. Se la popolazione e' molto variabile, ma puo' essere suddivisa in sot-

toinsiemi singolarmente piu' omogenei, si preferisce operare separatamente sui sottoinsiemi (con estrazioni casuali) e mediare poi i risultati (CAMPIONAMENTI PER STRATIFICAZIONE).

ESEMPIO

Vi sono 1000 persone di eta' compresa tra i 20 e i 21 anni residenti nella citta' X; sono interessato al loro peso medio e voglio limitare l'indagine ad un campione di 100 persone.

PROCEDIMENTO 1: scelgo a caso 100 persone, rilevo i dati e medio i pesi

PROCEDIMENTO 2: scelgo a caso 50 ragazzi e 50 ragazze, calcolo la media dei due gruppi e medio i risultati.

Che succede se la popolazione e' di 600 ragazze e 400 ragazzi? Nel primo caso il campione conterra', probabilmente, piu' femmine che maschi ed il valore ottenuto con il procedimento 1 sara' piu' attendibile di quello ricavato con il procedimento 2.

Se pero' so, fin dall'inizio, che l'universo e' formato da 600 ragazze e 400 ragazzi, posso considerare un campione stratificato di $60+40$ ottenendo un valore ancora piu' attendibile (si sta lavorando nell'ipotesi che i due insiemi siano singolarmente piu' omogenei della totalita').

Quando si utilizzano campioni casuali, l'inferenza statistica fornisce tutta una serie di strumenti che consentono di valutare con quale grado di fiducia le informazioni ricavate dal campione si possano riferire a tutta la popolazione. Viceversa, fissato il margine di errore che si e' disposti a tollerare, si potra' determinare qual e' la dimensione minima del campione necessario.

Tre esempi

Concludiamo con tre esempi (il 2' e il 3' sono presi da [14] pag. 53 e pag. 58) che diano un'idea dei problemi tipici dell'inferenza statistica e del tipo di approccio per risolverli; gli studenti degli ultimi anni della secondaria superiore troveranno in [4] (volume 3) una trattazione molto piu' esauriente.

ESEMPIO 1

Vorrei sapere quanti, tra i 6 vicini di casa, possiedono un televisore a colori.

Decido di chiedere a tre di loro e, dopo averli numerati da 1 a 6, lancio tre dadi ed intervisto le persone corrispondenti ai numeri ottenuti. Degli intervistati, due mi dicono di avere un apparecchio a colori, il terzo no.

Come stanno le cose in relazione a tutto il vicinato?

SOLUZIONE

L'universo statistico e' costituito da 6 unita'. Il campione, di dimensione 3, viene determinato attraverso una ESTRAZIONE BER-

NOULLIANA (o CON RIPETIZIONE): dopo l'estrazione, l'unita' statistica sorteggiata torna a far parte della popolazione e puo' essere nuovamente estratta (con un universo statistico di dimensioni tanto ridotte, questo modo di procedere appare ridicolo poiche' la stessa persona puo' essere intervistata due volte o addirittura tre).

Identifichiamo i vicini con 6 palline contenute in una scatola: i possessori di televisori a colori corrispondano a palline bianche, gli altri a palline nere.

Il problema diviene:

''Se una scatola contiene 6 palline tra bianche e nere e ne vengono estratte 3 con reimbussolamento ottenendone 2 bianche e 1 nera, che si puo' dire del contenuto della scatola?''

Indichiamo con x il numero di palline bianche presenti nella scatola e con Hx l'evento

Hx = {la scatola contiene x bianche e 6-x nere}.

Prima di eseguire le estrazioni, sappiamo solo che x puo' essere un qualsiasi intero da 0 a 6: l'assenza di informazione equivale all'ipotesi $P(Hx) = 1/7$ per ogni x.

Posto

A = {vengono estratte 2 bianche e 1 nera}

e'

$$P(A/Hx) = \binom{3}{2} \left(\frac{x}{6}\right)^2 \left(\frac{6-x}{6}\right) = \frac{x^2\,(6-x)}{36}$$

si tratta infatti della probabilita' di avere 2 successi (successo = estrazione di una bianca) su 3 prove in uno s.p.r. di parametro incognito $p = x/6$ (vedi 4.5). Giustamente $P(A/H_0) = P(A/H_6) = 0$ dato che A non puo' verificarsi se le palline sono tutte bianche o tutte nere.

Il problema consiste nel valutare P(Hx/A) in quanto sappiamo che A si e' verificato mentre non conosciamo x. Ci viene in aiuto la formula di Bayes (di cui si parlava nel problema n.7 del 2' capitolo); poiche' Hx $x = 0,1,..,6$ costituisce una partizione di Ω, e'

$$P(Hx/A) = \frac{P\,(A/H_x) \cdot P\,(Hx)}{\sum_{j=0}^{6} P(A/H_j) \cdot P(H_j)}$$

Calcoliamo il denominatore:

$$\sum_{j=0}^{6} P(A/Hj)\ P(Hj) = \sum_{j=0}^{6} \frac{j^2\,(6-j)}{36} \cdot \frac{1}{7} =$$

$$= \frac{1}{36} \cdot \frac{1}{7} [0 + 5 + 16 + 27 + 32 + 25] = \frac{5}{12}$$

quindi

$$P(Hx/A) = \frac{x^2(6-x)}{36} \cdot \frac{1}{7} \cdot \frac{12}{5} = \frac{x^2(6-x)}{105}$$

Per $x = 0,\ldots,6$ abbiamo:

x=	P(Hx/A) =
0	0
1	5/105 = 0.05
2	16/105 = 0.15
3	27/105 = 0.26
4	32/105 = 0.30
5	25/105 = 0.24
6	0

Il valore piu' alto si ottiene per $x = 4$ (come ci si aspettava) anche se la probabilita' associata non e' poi elevatissima. La tabella fornisce pero' informazioni piu' utili, ad esempio:

$P[3 \leqslant x \leqslant 5] = P(3) + P(4) + P(5) = 0.8$;

posso ora affermare, con un margine di sicurezza dell'80%, che il numero di vicini che possiedono un televisore a colori e' compreso tra 3 e 5 mentre prima dell'esperimento statistico tale probabilita' era pari a $3/7 = 43\%$.

ESEMPIO 2

Una fabbrica produce componenti elettronici in confezioni da 100 pezzi che provengono da due macchine M1 e M2 e precisamente: 60 pezzi da M1, 40 da M2.

A causa di un temporaneo malfunzionamento di M2, alcune confezioni contengono 40 circuiti difettosi. Ne controllo una verificando 5 pezzi scelti a caso: risultano perfettamente funzionanti.

Posso essere sicuro, al 95 %, della bonta' di tutti i cento pezzi della confezione in esame?

SOLUZIONE

Interessa calcolare la probabilita' che una confezione difettosa superi il controllo ovvero la probabilita' che, nelle 5 estrazioni, siano sempre usciti pezzi funzionanti quando la confezione ne contiene

40 difettosi.

La probabilita' richiesta e' (estrazioni SENZA reimbussolamento):

$$P = \frac{60}{100} \cdot \frac{59}{99} \cdot \frac{58}{98} \cdot \frac{57}{97} \cdot \frac{56}{96} = 0.07254.. \cong 7.2\ \%$$

P e' la ''probabilita' di sbagliarsi'' cioe' la probabilita' di considerare valida una confezione difettosa. Poiche' 7.2% > 5%, non posso essere sicuro al 95% (ma solo al 92.8%).

Se avessi eseguito le 5 estrazioni CON reimbussolamento, la probabilita' di ''errore'' sarebbe stata:

$$P' = \left(\frac{60}{100}\right)^5 \cong 0.078 = 7.8\ \% .$$

Supponendo di eseguire estrazioni con reimbussolamento, quanti pezzi devo controllare (come minimo) per essere sicuro della bonta' della confezione al 95%? E al 99%?

Sia ''n'' il numero di circuiti da controllare; ''n'' e' il piu' piccolo intero per cui:

$(0.6)^n \leqslant 0.05$ per il 95%

e $(0.6)^n \leqslant 0.01$ per il 99%

n	$(0.6)^n$
1	0.60000
2	0.36000
3	0.21600
4	0.12960
5	0.07776
6	0.04666 ***
7	0.02799
8	0.01680
9	0.01008
10	0.00605 ***

Quindi 6 pezzi per essere sicuri al 95% e 10 pezzi per il 99%.

Il risultato si ottiene piu' rapidamente passando ai logaritmi. Deve essere:

$\log (0.6)^n \leqslant \log 0.05$ per il 95 %

e

$\log (0.6)^n \leqslant \log 0.01$ per il 99 %

da cui, rispettivamente:

$n \cdot \log 0.6 \leqslant \log 0.05$

e n · log 0.6 $\leqslant$ log 0.01
poiche' log 0.6 $<$ 0 e'

$$n \geqslant \frac{\log 0.05}{\log 0.6} \cong 5.864$$

per il 95% : il minimo intero e' 6; mentre

$$n \geqslant \frac{\log 0.01}{\log 0.6} \cong 9.0151$$

per il 99% : il minimo intero e' 10.

ESEMPIO 3

Si vuol sapere qual e' stato il numero medio di giorni di degenza dei pazienti ricoverati presso un certo ospedale nel 1986.

Marinella e Sergio, incaricati dell'indagine, procedono in modo diverso: Marinella, procuratasi un elenco delle persone ricoverate, ne estrae un campione su cui lavorare; Sergio sceglie a caso un certo numero di date del 1986 e prende in considerazione le persone che risultano ricoverate in quei giorni. Chi dei due otterra' risultati piu' attendibili?

SOLUZIONE

Campione distorto

Il campione utilizzato da Sergio risulta essere un CAMPIONE DISTORTO; a parita' di dimensioni, il valore da lui calcolato sara', probabilmente, piu' elevato di quello valutato da Marinella.

Con il metodo di Sergio infatti, le persone costrette a lunghe degenze hanno maggiore probabilita' di entrare a far parte del campione contribuendo ad aumentare il valore medio.

Consideriamo un "caso limite" in cui, di 100 pazienti, 50 siano stati ricoverati dal 1' gennaio al 10 gennaio e i rimanenti 50 dall' 11 gennaio al 31 dicembre (una situazione davvero poco invidiabile).

Mentre Marinella estrarra' un campione che, in linea di massima, conterra' meta' elementi del primo gruppo e meta' del secondo, quello di Sergio sara' costituito quasi esclusivamente da persone del secondo gruppo, poiche' le date scelte casualmente cadranno quasi tutte oltre il 10 gennaio.

Nella realta', il carattere in esame risulta molto variabile e converrebbe far uso di campioni stratificati, studiando separatamente i vari reparti in diversi periodi dell'anno.

CAPITOLO 4

CALCOLO COMBINATORIO E DISTRIBUZIONE BINOMIALE

Il calcolo combinatorio si occupa della formazione e del conteggio dei raggruppamenti che si possono realizzare utilizzando gli elementi di un insieme finito (o di piu' insiemi finiti).

Poiche' si tratta di una parte affrontata in quasi tutti i libri di testo, ci limiteremo agli argomenti principali. Considereremo un insieme A formato da n elementi.

4.1 DISPOSIZIONI SEMPLICI

Si definiscono "disposizioni semplici di classe k" ($1 \leq k \leq n$) tutti i raggruppamenti di k elementi distinti di A tali che ogni raggruppamento differisca dagli altri

- per la natura degli elementi (cioe' per avere almeno un elemento diverso)

  oppure
- per l'ordine degli elementi (cioe' per avere gli elementi disposti diversamente).

Es.1 Sia

$A = \{a,b,c,d\}$

quindi $n = 4$; costruiamo le disposizioni semplici di classe 3 ($k = 3$):

abc	abd	acd	bcd
acb	adb	adc	bdc
bac	bad	cad	cbd
bca	bda	cda	cdb
cab	dab	dac	dbc
cba	dba	dca	dcb

le terne in colonne diverse differiscono per almeno un elemento, quelle nella medesima colonna differiscono per l'ordine.

Il numero delle disposizioni semplici di n oggetti della classe k si indica con $D_{n,k}$ ed e'

$$D_{n,k} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \qquad (*)$$

cosi' $D_{4,3} = 4 \cdot 3 \cdot 2 = 24$. Si noti che il numero di fattori e' k per cui $D_{n,k}$ e' dato dal prodotto di k fattori naturali decrescenti a partire da n ($D_{5,2} = 5 \cdot 4 = 20$; $D_{5,3} = 5 \cdot 4 \cdot 3 = 60$; $D_{7,3} = 7 \cdot 6 \cdot 5 = 210$).

Dimostrazione della (*). Supponiamo di aver elencato le disposizioni di classe k-1; per ottenere quelle di classe k bastera' aggiungere ad ogni raggruppamento uno degli $n-(k-1) = n-k+1$ elementi non ancora utlizzati. Dunque $D_{n,k} = (n-k+1) \cdot D_{n,k-1}$; poiche' $D_{n,1} = n$ avremo:

$$\begin{aligned} D_{n,1} &= n \\ D_{n,2} &= (n-2+1) \cdot D_{n,1} = (n-1) \cdot D_{n,1} \\ D_{n,2} &= (n-3+1) \cdot D_{n,2} = (n-2) \cdot D_{n,2} \\ &\ldots \\ D_{n,k} &= (n-k+1) \cdot D_{n,k-1} \ ; \end{aligned}$$

uguagliando il prodotto dei primi membri con quello dei secondi membri e sopprimendo i fattori comuni si ottiene la (*).

4.2 DISPOSIZIONI CON RIPETIZIONE

La definizione e' analoga a quella data al punto precedente con la differenza che lo stesso oggetto puo' apparire nel raggruppamento piu' volte cioe' da 0 a k volte e che k puo' essere maggiore di n.

Es.2 Sia

$A = \{a,b,c,d\}$

quindi $n = 4$. Le disposizioni con ripetizione di classe 2 ($k = 2$) sono:

aa	ab	ac	ad
ba	bb	bc	bd
ca	cb	cc	cd
da	db	dc	dd

Es.3 Sia $A = \{x,y\}$ quindi $n = 2$. Le disposizioni con ripetizione della classe 3 ($k = 3$) sono:

xxx xxy xyx xyy

yxx yxy yyx yyy

(in questo esempio e' $k>n$).

Il numero delle disposizioni di n oggetto della classe k si indica con $D'_{n,k}$ e si ha:

$$D'_{n,k} = n^k.$$

In effetti, supponendo di aver elencato le disposizioni con ripetizione di classe k-1, quelle di classe k si otterranno aggiungendo a ciascuna delle precedenti uno degli n oggetti. Quindi

$$D'_{n,k} = n \cdot D'_{n,k-1} ;$$

poiche'

$$D'_{n,1} = n,$$

avremo

$$D'_{n,2} = n \cdot D'_{n,1} = n \cdot n = n^2, \; D'_{n,3} = n \cdot D'_{n,2} = n \cdot n^2 = n^3$$

e, in generale,

$$D'_{n,k} = n^k.$$

Es.4 Quante sono le diverse colonne della schedina del totocalcio? Sono tante quante le disposizioni con ripetizione di 3 oggetti (1,X,2) delle classe 13: $D'_{3,13} = 3^{13} = 1594323$.

Es.5 Una scatola contiene 3 palline di colore differente. Si eseguono 2 estrazioni CON reimbussolamento. Calcolare la probabilita' dell'evento

A = {le palline estratte sono di colore diverso}.

Soluzione 1. Qualunque sia l'esito della prima estrazione, l'evento A si verifica se alla seconda si ottiene un colore diverso dalla prima. Vi sono 3 casi possibili di cui 2 favorevoli: P(A) = 2/3.

Soluzione 2. I casi possibili sono le disposizioni con ripetizione (reimbussolamento) di 3 oggetti (le palline) a gruppi di 2. Il loro numero e' $D'_{3,2} = 9$. I casi favorevoli sono le disposizioni semplici (colore diverso) in numero di $D_{3,2} = 6$. Allora P(A) = 6/9 = 2/3.

4.3 PERMUTAZIONI

Le disposizioni semplici di n oggetti della classe n vengono dette

permutazioni; il loro numero si indica con P_n. Per definizione e'

$$P_n = D_{n,n} = n(n-1)(n-2) \dots (n-n+1) = n(n-1)(n-2) \dots 3 \cdot 2 \cdot 1 = n!$$

dove si e' indicato con n! (''enne fattoriale'') il prodotto degli interi da 1 a n.

Dato che nelle permutazioni compaiono tutti gli oggetti, una permutazione differisce da un'altra per l'ordine.

Es.6 Sia

$A = \{a,b,c\}$,

le permutazioni sono

abc bac cab acb bca cba ,

$P_3 = 3! = 3 \cdot 2 \cdot 1 = 6$.

4.4 ANAGRAMMI

Se due diverse parole sono formate dalle stesse lettere alfabetiche, si dice che una e' un anagramma dell'altra (es. CASO CAOS COSA; TORTA TROTA ROTTA RATTO). Ci proponiamo di calcolare il numero di anagrammi distinti (anche senza significato) di una parola di n lettere (compresa la parola stessa).

Se le lettere sono tutte diverse, la risposta e' $P_n = n!$ poiche' gli anagrammi coincidono con le permutazioni delle n lettere. Cosi' gli anagrammi della parola ALT sono 6:

ALT, ATL, LAT, LTA, TAL, TLA.

Se una lettera compare piu' volte, il numero di anagrammi e' inferiore a n! in quanto andranno identificate quelle parole che si ottengono dalle permutazioni della lettera ripetuta e questo per ogni posizione fissata nell'ambito della parola. Cosi' in AMACA (n=6), uno scambio tra la prima e la seconda ''A'' non origina una nuova parola: poiche' le permutazioni delle tre ''A'' sono in numero di $3! = 6$, il numero di anagrammi distinti sara' $6!/3! = 120$.

Il ragionamento si estende in modo analogo al caso di piu' lettere ripetute: se in una parola di n lettere

una lettera e' ripetuta k_1 volte

un'altra lettera e' ripetuta k_2 volte

una terza lettera e' ripetuta k_3 volte

...

il numero di anagrammi e' $n!/(k_1! \cdot k_2! \cdot k_3! \cdot \dots)$.

Es.7 Nella parola ORO di 3 lettere, la lettera O compare 2 volte; il numero di anagrammi e' $3!/2! = 3$ e sono ROO, ORO, OOR.

Es.8 Nelle parola PAPA di 4 lettere

- la lettera P compare 2 volte
- la lettera A compare 2 volte;

il numero di anagrammi e' $4!/(2!\cdot 2!) = 6$: AAPP, APAP, APPA, PPAA, PAPA, PAAP.

Es. 9 ''MATEMATICA'': $n = 10$; due M, tre A, due T. Il numero di anagrammi e' $10!/(2!\cdot 3!\cdot 2!) = 151200$.

Di particolare interesse e' il caso di una parola di n lettere in cui una stessa lettera compare k volte e una seconda lettera le rimanenti n-k volte (come in AABB $n = 5$, $k = 3$). Il numero di anagrammi e':

$$\frac{n!}{k!\,(n-k)!}$$

valore che viene detto COEFFICIENTE BINOMIALE e che si indica con il simbolo $\binom{n}{k}$ da leggersi ''n su k'' o ''n sopra k''. Se $k = 0$ o $k = n$, le n lettere sono tutte uguali e il numero di anagrammi si riduce a 1; ponendo $0! = 1$, la formula vale per $0 \leq k \leq n$.

Nello sviluppo di $(a+b)^n$, il coefficiente di $a^k b^{n-k}$ risulta pari al numero di anagrammi di una parola formata da k lettere ''a'' e n-k lettere ''b''; ad esempio:

$$\begin{aligned}(a+b)^3 &= (a+b)(a+b)(a+b) = (aa+ab+ba+bb)(a+b) = \\ &= aaa+aba+baa+bba+aab+abb+bab+bbb = \\ &= a^3 + 3a^2b + 3ab^2 + b^3;\end{aligned}$$

in generale:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k};$$

da qui la denominazione di coefficiente binomiale attribuita a $\binom{n}{k}$.

4.5 DISTRIBUZIONE BINOMIALE

In uno schema delle prove ripetute di parametro p $(0 < p < 1)$, la probabilita' che si verifichino k successi in n prove $(0 \leq k \leq n)$ e' data da:

$$b(k,n;p) = \binom{n}{k} p^k q^{n-k} \qquad \text{dove } q = 1-p.$$

In effetti, l'evento

$A = \{k \text{ successi in } n \text{ prove}\}$

si verifica se si realizza una qualsiasi sequenza del tipo SFFSS...FFS

dove S (successo) compare k volte e F (fallimento) n-k volte. Ogni sequenza ha probabilita' $p^k q^{n-k}$, mentre il numero delle diverse sequenze e' dato dagli anagrammi della ''parola'' SSS..SFFF..F ovvero $\binom{n}{k}$.

Per l'additivita':

$$P(A) = \binom{n}{k} p^k q^{n-k}.$$

Es.10 Un dado regolare viene lanciato 10 volte; calcolare la probabilita' che la faccia ''6'' compaia esattamente 3 volte.

Soluzione. Detto ''successo'' l'uscita del 6, e' $p = 1/6$, $q = 5/6$, $n = 10$, $k = 3$. La probabilita' richiesta e'

$$\binom{10}{3} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^7 \simeq 0.155$$

Es.11 Una moneta regolare viene lanciata 5 volte; calcolare la probabilita' che la faccia testa compaia 0 volte, 1 volta, 2 volte,..., 5 volte.

Soluzione. Detto ''successo'' l'uscita ''T'', e' $p = q = 1/2$ e $n = 5$. Per $k = 0,1,..,5$ sara'

$$P(k) = \binom{5}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{5-k} = \binom{5}{k} \cdot \left(\frac{1}{2}\right)^5$$

Calcolando:

$P(0) = 1/32$; $P(1) = 5/32$; $P(2) = 10/32$; $P(3) = 10/32$; $P(4) = 5/32$; $P(5) = 1/32$.

Si noti che $P(0) + P(1) + \ldots + P(5) = 1$ e, in generale,

$$\sum_{k=0}^{n} b(k,n;p) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1.$$

4.6 COMBINAZIONI SEMPLICI

Dato l'insieme A di n elementi, si definiscono COMBINAZIONI SEMPLICI di classe k i raggruppamenti di k elementi distinti di A tali che ogni raggruppamento differisca dagli altri per la natura degli elementi.

Es.12 Sia

$A = \{a,b,c,d,e\}$,

quindi $n = 5$. Costruiamo le combinazioni di classe 3:

abc abd abe acd ace

ade bcd bce bde cde.

Due qualsiasi combinazioni differiscono per avere almeno una lettera diversa (''abc'' e ''acb'' costituiscono due disposizioni distinte ma la stessa combinazione).

Il numero delle combinazioni semplici di n oggetti della classe k si indica con $C_{n,k}$ (quindi $C_{5,3} = 10$).

Vale la seguente uguaglianza:

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n(n-1) \cdot \ldots \cdot (n-k+1)}{k!} \qquad (**)$$

in effetti, permutando in tutti i modi possibili i k oggetti di ciascuna combinazione si ottengono le disposizioni semplici per cui $C_{n,k} \cdot P_k = D_{n,k}$ (si consideri l'insieme D delle disposizioni semplici e si introduca la relazione di equivalenza $d \sim d'$ se d e d' sono formate dagli stessi oggetti. Ogni classe di equivalenza contiene esattamente P_k disposizioni e l'insieme quoziente sara' formato da $D_{n,k}/P_k$ classi di equivalenza).

Applicando la (**) si ha:

$$C_{5,3} = \frac{D_{5,3}}{P_3} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

$$C_{7,5} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 21$$

Es.13 Qual e' il numero di diagonali di un poligono non intrecciato di n lati?

Soluzione. Dopo aver numerato i vertici da 1 a n, le diagonali saranno tante quante le coppie non ordinate del tipo $\{h,k\}$ con $h \neq k$ eccettuate le n coppie che identificano i lati. Il numero di diagonali e' $C_{n,2} - n = n(n-3)/2$.

Moltiplicando per (n-k)! entrambi i termini della frazione nella (**) si ottiene:

$$C_{n,k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Proprieta' dei coefficienti binomiali

I coefficienti binomiali godono di diverse proprieta', tra cui:

$$\binom{n}{k} = \binom{n}{n-k} \;;$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \;;$$

$$\binom{n}{k+1} = \binom{n}{k} \cdot \frac{n-k}{k+1}$$

La prima, di verifica immediata, si spiega intuitivamente: per ogni combinazione di k elementi ne rimane una costituita dagli n-k non utilizzati per cui $C_{n,k} = C_{n,n-k}$.

Per quanto riguarda la seconda abbiamo:

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{k!\,(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} =$$

$$= \frac{(n-1)!}{(k-1)!(n-1-k)!} \left[\frac{1}{k} + \frac{1}{n-k} \right] =$$

$$= \frac{(n-1)!}{(k-1)!(n-1-k)!} \cdot \frac{n}{k(n-k)} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

dove si e' tenuto conto che $m! = m(m-1)!$.

I $C_{n,k}$ anagrammi della parola AA..ABB..B formata da k lettere A e n-k lettere B iniziano con A oppure con B.

Quelli che iniziano con A sono tanti quanti gli anagrammi di una parola di n-1 lettere di cui k-1 sono A; il loro numero e' $C_{n-1,k-1}$.

Quelli che iniziano con B sono tanti quanti gli anagrammi di una parola di n-1 lettere di cui k sono A; il loro numero e' $C_{n-1,k}$.

Quindi $C_{n,k} = C_{n-1,k-1} + C_{n-1,k}$.

Triangolo di Pascal

Della seconda proprieta' si fa uso nella costruzione del cosiddetto ''triangolo di Pascal'' (o ''triangolo di Tartaglia'').

Di facile verifica la terza proprieta' che fornisce un'utilissima relazione ricorrente per il calcolo dei termini della distribuzione binomiale:

$$\binom{n}{k+1} = \frac{n!}{(k+1)!(n-k-1)!} = \frac{n!\,(n-k)}{(k+1)\cdot k!\cdot(n-k)!} = \binom{n}{k} \cdot \frac{n-k}{k+1}$$

da cui

$$b(k+1,n;p) = b(k,n;p) \cdot \frac{(n-k)p}{(k+1)q}$$

Il programma CALCOLO COMBINATORIO

Questa relazione e' stata utilizzata nel programma BINOMIALE.

Il programma CALCOLO COMBINATORIO consente il calcolo del numero dei vari raggruppamenti e l'elencazione degli stessi.

Anche per modesti valori di n e k, il numero di raggruppamenti puo' essere molto elevato (ad esempio, le permutazioni di 10 oggetti sono 3628800), per cui conviene evitare la stampa di elenchi molto lunghi (soprattutto su carta) che, inoltre, richiedono tempi

non indifferenti per essere elaborati. Per le routine di generazione, si veda [10] e [13].

Esempi di Output del programma CALCOLO COMBINATORIO:

1. Disposizioni semplici: n = 4; k = 2
   D (4,2) = 12
ab ac ad ba bc bd ca cb cd da db dc

2. Disposizioni con ripetizione: n = 2; k = 3
   D'(2,3) = 8
aaa aab aba abb baa bab bba bbb

3. Permutazioni: n = 3
   P (3) = 6
abc acb bac bca cab cba

4. Combinazioni semplici: n = 5; k = 3
   C(5,3) = 10

abc abd abe acd ace ade bcd bce bde cde

BIBLIOGRAFIA

[1] BETTAZZI, G. e MONTI, A. ''Calcolo delle probabilita' e statistica metodologica'' Bologna: Zanichelli, 1984

[2] DE FINETTI, B. ''Teoria delle probabilita'''(2 voll.) Torino: Einaudi, 1970

[3] DE MICHELE, F. e ROSA-CLOT, M. ''Probabilita' statistica e termodinamica'' Milano: Jackson, 1986

[4] GAMBOTTO MANZONE, A. ''Matematica per ragionieri programmatori'' (voll. 1 e 2) Bresso: Tramontana, 1985

[5] GLAYMANN, M. e VARGA, T. ''La probabilita' nella scuola dell'obbligo'' Roma: Armando, 1979

[6] HAUT, H. ''Programmi di matematica e statistica in Basic'' Milano: Jackson, 1982

[7] IAZEOLLA, G. ''Introduzione alla simulazione discreta'' Torino: Boringhieri, 1978

[8] LOMBARDO RADICE, L. e MANCINI PROIA, L. ''Il metodo matematico'' (voll. 1 e 2) Milano: Principato, 1982

[9] MAISEL, H. e GNUGNOLI, G. ''Simulation of Discrete Stochastic Systems'' Chicago: S.R.A., 1972

[10] PAGE, E.S. e WILSON, L.B. ''La combinatoria computazionale'' Padova: Muzzio, 1980

[11] PAMPALLONA, U. e RAGUSA GILLI, L. ''Che cos'e' la statistica'' Torino: Loescher, 1979

[12] PICCINATO, L. e PINTACUDA, N. ''Probabilita' e statistica'' a cura del C.N.R., 1985

[13] PINTACUDA, N. ''Algoritmi elementari'' Padova: Muzzio, 1986

[14] PINTACUDA, N. ''Insegnare la probabilita''' Padova: Muzzio, 1981

[15] PINTACUDA, N. ''Primo corso di probabilita''' Padova: Muzzio, 1983

[16] RICCI, F. ''Statistica'' Bologna: Zanichelli, 1975

[17] ''School Mathematics Project'' (vol. 5) Bologna: Zanichelli, 1976

[18] SPERANZA, F. - MEDICI CAFFARRA, D. - QUATTROCCHI, P. ''Insegnare la matematica nella scuola elementare'' Bologna: Zanichelli, 1986

[19] SPERANZA, F. e ROSSI DELL'ACQUA, A. ''Il linguaggio della matematica'' Bologna: Zanichelli, 1979

Laboratorio Scuola

P.F. Tramontano
ALGEBRA
Teoria, esercizi e simulazioni
Il software interviene a completare il testo laddove il calcolatore può arricchire in modo corretto la materia rendendola didatticamente più valida e fornendo materiale per integrare i corsi istituzionali di Algebra Elementare.

120 pagine
Libro + disco
versione software per:

MS-DOS
Cod. SD2921 L. 50.000

M20 ●
Cod. 292DOL L. 50.000

C64 ●
Cod. 292DCM L. 50.000

● a richiesta

M. Beretta
IDRAULICA
Teoria, esercizi e simulazioni
Con il binomio libro-software vengono ripresi alcuni aspetti classici della materia sia dal punto di vista teorico sia dal punto di vista applicativo attraverso le simulazioni consentite dal computer.

144 pagine
Libro + cassetta/disco
versione software per:

C64
Cod. SD296 L. 40.000

MS-DOS
Cod. SD296I L. 50.000

Apple ●
Cod. 296DAP L. 50.000

C64 ●
Cod. 296DCM L. 50.000

● a richiesta

P.F. Tramontano, B. Rinaldi
ANALISI
Teoria, esercizi e simulazioni
Il testo si rivolge in particolar modo a studenti delle scuole superiori ed il software allegato interviene a completare il testo laddove il calcolatore può arricchire in modo concreto la materia rendendola didatticamente più valida e fornendo materiale per integrare i corsi istituzionali della disciplina.

224 pagine
Libro + disco
versione software per:

MS-DOS
Cod. SD293I L. 50.000

M20 ●
Cod. 293DOL L. 50.000

C64 ●
Cod. 293DCM L. 50.000

● a richiesta

A. Conti, M. Rosa - Clot
MECCANICA
Teoria, esercizi e simulazioni
Il software allegato al testo affronta in 30 problemi i principi base della cinematica e della dinamica, coprendo l'intero programma ministeriale di Fisica.

300 pagine
Libro + cassetta/disco
versione software per:

C64
Cod. SD290 L. 40.000

MS-DOS
Cod. SD290I L. 50.000

Apple ●
Cod. 290DAP L. 50.000

C64 ●
Cod. 290DCM L. 50.000

● a richiesta

A. Martinelli, M. Persico

CHIMICA

Teoria, esercizi e simulazioni

La teoria molecolare e la stechiometria possono essere apprese attraverso esercizi e simulazioni mediante il software allegato al libro.

192 pagine

Libro + cassetta/disco versione software per:

C64
Cod. SD295 L. 40.000

MS-DOS
Cod. SD295I L. 50.000

C64 ●
Cod. 295DCM L. 50.000

● a richiesta

L. Fronzoni, S. Faetti

COMPORTAMENTO DEI SISTEMI COMPLESSI: ORDINE E CAOS

Teoria, esercizi e simulazioni

Il disco, o la cassetta, insieme al libro hanno come scopo la descrizione logico-matematica dei sistemi complessi. Vengono discussi alcuni modelli per la descrizione delle instabilità ecodinamiche, laser, dinamica delle popolazioni. Ciascun sistema presentato viene analizzato secondo le linee seguenti:

- descrizione
- parametrizzazione
- comportamento dinamico (simulazione attraverso il calcolatore)

160 pagine

Libro + cassetta/disco versione software per:

C64
Cod. SD297 L. 40.000

MS-DOS
Cod. SD297I L. 50.000

C64 ●
Cod. 297DCM L. 50.000

● a richiesta

E. Boni, S. Schacheri

FISICA ATOMICA

Teoria, esercizi e simulazioni

Il dischetto insieme al libro si propone di presentare alcuni tra gli esperimenti più importanti che hanno portato alla "costruzione", oggi accettata, del modello di atomo

168 pagine

Libro + cassetta/disco versione software per:

C64
Cod. SD300 L. 40.000

MS-DOS
Cod. SD300I L. 50.000

Apple ●
Cod. 300DAP L. 50.000

C64 ●
Cod. 300DCM L. 50.000

● a richiesta

L. Danusso, S. Schacheri

CAMPI ELETTRICI E MAGNETICI

Teoria, esercizi e simulazioni

Il calcolatore permette la simulazione di esperimenti altrimenti difficili da elaborare e preparare nella pratica di laboratorio, per una migliore comprensione di questa materia. Gli argomenti trattati coprono il programma ministeriale di elettromagnetismo per le medie superiori.

96 pagine

Libro + cassetta/disco versione software per:

C64
Cod. SD301 L. 40.000

MS-DOS
Cod. SD301I L. 50.000

Apple ●
Cod. 301DAP L. 50.000

C64 ●
Cod. 301DCM L. 50.000

● a richiesta

Didattica e nuove tecnologie

C. Bocchetti, L. Accomazzi
L'INSEGNANTE E IL CALCOLATORE
Come e perché il computer nella scuola
Cod. SF226
176 pagine **L. 16.500**
La trasformazione della scuola in seguito alla sua progressiva informatizzazione, lo sviluppo in parallelo di nuovi strumenti e tecniche di insegnamento, costituiscono il nucleo centrale di questo libro.

C. Tirittico, R. Traversini
IL CAI
Un manuale per l'uso dell'elaboratore nella didattica
Cod. SF227
128 pagine **L. 13.000**
Partendo da un'analisi dettagliata delle tecniche e degli strumenti per l'uso dell'elaboratore nella didattica, il testo fornisce metodi e suggerimenti per affrontare la scelta e il dimensionamento di sistemi per l'insegnamento con l'ausilio del Computer (CAI).

G. Mauri
COMUNICARE CON IL COMPUTER: I LINGUAGGI
Cod. SD284
148 pagine **L. 15.000**
Il libro fornisce un utile strumento che consente all'insegnante la formulazione di ipotesi su cui operare in un contesto non rigidamente predeterminato.

C. Poma
IL COMPUTER NELLA SCUOLA MEDIA
Cod. 508A
200 pagine **L. 18.000**
Un libro adatto a chiunque si sta avvicinando per la prima volta al mondo dell'informatica e della programmazione; indicato in modo specifico per studenti e insegnanti delle scuole medie.

G. Tonfoni
LA COMUNICAZIONE CAMBIATA
Cod. 538P
116 pagine **L. 10.500**
Un libro per scoprire, analizzare, potenziare le capacità comunicative di ciascuno grazie ai nuovi strumenti della tecnologia.

L. Ortolani, O. Cepelli
VALUTAZIONE SCOLASTICA CON L'AIUTO DEL COMPUTER
Cod. SD232
136 pagine **L. 34.000**
Supporto DISCO PER APPLE
Elementi fondamentali della statistica su cui fondare un metodo di valutazione scolastica semplice ed oggettivamente valido.

A.I.C.A.
SOFTWARE DIDATTICO
Cod. 804C
408 pagine **L. 45.000**
Di ogni programma sono descritti gli obiettivi, gli elementi caratterizzanti dal punto di vista didattico e degli allievi a cui si rivolge, gli strumenti informatici su cui funziona e la documentazione di cui è corredato.

E. Tonti
DIDATTICA CON IL PERSONAL COMPUTER
Cod. 400A
156 pagine **L. 24.000**
Un libro per docenti orientati all'impiego del calcolatore in classe, con programmi di immediato utilizzo e spiegazioni sulla loro funzione didattica.

"Ogni scatola di detersivo contiene una figurina che trova posto su un album che ne ospita 100. Quante scatole di detersivo si dovranno acquistare mediamente per completare la raccolta?". Ecco un bel problema di calcolo delle probabilità: un solo dato, una situazione abbastanza realistica, una soluzione non banale.
È questo uno dei 17 problemi proposti e risolti nel secondo capitolo di questo libro; la maggior parte dei programmi associati agli esercizi realizzano delle simulazioni il cui scopo è quello di verificare "sul campo" la validità dei risultati ottenuti teoricamente. Ecco dunque il calcolatore gestire gli acquisti di una nutrita schiera di collezionisti e sfornare, in pochi minuti, risultati che richiederebbero intere giornate di simulazione "manuale" (a proposito, per un album di 100 figurine, si dovranno acquistare mediamente 519 scatole di detersivo).
Gli esercizi sono preceduti da un capitolo dedicato alla teoria del calcolo delle probabilità; di statistica si parla, a livello molto elementare, nel terzo capitolo. Chi è interessato ad approfondire questa parte troverà, nella stessa collana, il volume "Probabilità, Statistica e Termodinamica" ulteriore materiale e il relativo software.

GRUPPO EDITORIALE JACKSON

ISBN 88-7056-790-7

L. 50.000

Cod. SD579

9 788870 567908